



Lehrstuhl für Verkehrstechnik
Technische Universität München
Univ.-Prof. Dr.-Ing. Fritz Busch

Schriftenreihe Heft 9

Axel Leonhardt

Ein Instanzbasiertes Lernverfahren zur
Prognose von Verkehrskenngrößen unter
Nutzung Räumlich-Zeitlicher Verkehrsmuster

München 2008

Die Deutsche Bibliothek – CIP Einheitsaufnahme

Leonhardt, Axel:

Ein Instanzbasiertes Lernverfahren zur Prognose von
Verkehrskenngrößen unter Nutzung Räumlich-Zeitlicher
Verkehrsmuster

Hrsg.: Fritz Busch, München, 2009

Schriftenreihe des Lehrstuhls für Verkehrstechnik
der Technischen Universität München, Heft 9

Zugl.: München, Techn. Univ., Diss., 2008

ISBN 978-3-937631-09-7

Copyright ©

Lehrstuhl für Verkehrstechnik der Technischen Universität München 2009

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder
vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der
Übersetzung, vorbehalten.

Druck: TypeSet GmbH, Ismaning

ISBN 978-3-937631-09-7

ISSN 1612-9431

Lehrstuhl für Verkehrstechnik · Institut für Verkehrswesen

Technische Universität München · 80333 München

Telefon: 089 / 289 – 22438 · Telefax: 089 / 289 – 22333 · E-Mail: info@vt.bv.tum.de

www.vt.bv.tum.de

Bei der vorliegenden Veröffentlichung handelt es sich um die Dissertation von
Herrn Dr.-Ing. Axel Leonhardt

Vorsitzender: Univ.-Prof. Dr.-Ing. Matthäus Schilcher, TU München

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Fritz Busch, TU München
2. Prof. Michael G. H. Bell, PhD,
Imperial College of Science, United Kingdom

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Verkehrstechnik der Technischen Universität München.

Mein besonderer Dank gilt meinem Doktorvater Herrn Professor Fritz Busch für die Betreuung der Arbeit. Er hat mir gleichermaßen viele Freiheiten gelassen und mit seinen wertvollen Anregungen – und immer wieder faszinierendem Blick für das Wesentliche – entscheidend zum Gelingen der Arbeit beigetragen. Bei Herrn Professor Michael G.H. Bell bedanke ich mich für die freundliche Übernahme des Koreferats. Es freut mich, dass die vor ein paar Jahren begonnene Zusammenarbeit thematisch mit der Dissertation fortgesetzt werden konnte.

Allen Kollegen, die auf verschiedene Weisen zum Gelingen und manchmal auch zur Ablenkung von dieser Arbeit beigetragen haben, danke ich ebenfalls. Insbesondere bedanke ich mich für die fachlichen Diskussionen, das Korrekturlesen und das freundschaftliche Umfeld an unserem Lehrstuhl.

Axel Leonhardt

München, im Dezember 2008

Kurzfassung

Gegenstand dieser Dissertation sind Entwicklung und Anwendung eines Instanzbasierten Lernverfahrens zur Prognose von Reisezeiten und lokalen Verkehrskenngrößen als Basis für dynamische Informations- und Steuerungsverfahren im Verkehrsmanagement.

Grundlage des Verfahrens bilden kontinuierlich erhobene lokale Verkehrskenngrößen, unregelmäßig beobachtete Reisezeiten, sowie Kalendermerkmale (Tagesklasse und Tageszeit). Die Verkehrskenngrößen werden unter Berücksichtigung der Geometrie des Straßennetzes räumlich-zeitlich strukturiert und für den späteren Zugriff archiviert.

Im Sinne des Regressionsgedankens werden für die zu prognostizierenden abhängigen Variablen ("Zielgrößen", Reisezeiten oder lokale Kenngrößen) relevante unabhängige Variablen ("Merkmale", lokale Kenngrößen und Kalendermerkmale) identifiziert und als zur Zielgröße gehöriges Muster betrachtet. Die Auswahl und die Gewichtung der verkehrlichen Merkmale basiert dabei auf räumlichen und zeitlichen Abständen zur Zielgröße. Die Kalendermerkmale werden relativ zu den Verkehrsmerkmalen und abhängig vom Prognosehorizont gewichtet. Bei einer Prognoseanfrage wird die Ähnlichkeit des aktuellen Musters zu jedem archivierten historischen Muster berechnet, für das eine Beobachtung der Zielgröße vorliegt. Aus den zu den ähnlichsten Mustern korrespondierenden Zielgrößen wird die Prognose als gewichteter Mittelwert berechnet.

Das Verfahren wurde implementiert und anhand von realen Daten aus Graz (Belegungsgrade und Taxipositionsmeldungen) und München (Verkehrsstärken und Reisezeiten aus Fahrzeugwiedererkennung) kalibriert und bewertet. Das Verfahren besitzt sechs wählbare Parameter zur Gewichtung der Merkmale im Mustererkennungsprozess und zur Approximation der Prognose. Eine Analyse zeigt die Sensitivität der Prognoseergebnisse in Abhängigkeit der in Wechselwirkung stehenden Parameter. Ein Genetischer Algorithmus wird zur automatischen und gleichzeitigen Kalibrierung der Parameter eingesetzt.

Das entwickelte Verfahren wird unter verschiedenen Randbedingungen getestet und bewertet und mit einfachen Verfahren (kalenderbasierte Prognose, Trivialprognose), sowie mit künstlichen Neuronalen Netzen verglichen. Sowohl bei der Prognose des Belegungsgrades, als auch bei der Prognose der Reisezeit sind die Ergebnisse des Instanzbasierten Lernverfahrens besser als die der einfachen Verfahren und vergleichbar mit denen der künstlichen Neuronalen Netze.

Wesentliche Vorzüge des Instanzbasierten Lernverfahrens sind die Adaptivität durch eine automatisch mitwachsende Wissensbasis und die Robustheit gegenüber fehlenden Eingangswerten. Darüber hinaus können neben Erwartungswerten auch Indikatoren für die Zuverlässigkeit der Prognose und Verteilungsdichtefunktionen der Zielgröße geschätzt werden.

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Kontext und Problemstellung	1
1.2	Anforderungen an das Verfahren.....	2
1.3	Lösungsansatz und methodischer Aufbau	3
2	Verkehrsdatenerfassung und -aufbereitung	7
2.1	Erfassung und Aufbereitung lokaler Verkehrsdaten	7
2.1.1	Lokale Verkehrsdatenerfassung	7
2.1.2	Beschreibung des lokalen Verkehrszustands.....	8
2.1.3	Aufbereitung lokaler Verkehrsdaten	9
2.2	Erfassung und Aufbereitung der Reisezeit	12
2.2.1	Grundlagen.....	13
2.2.2	Erhebung von Reisezeiten mittels Fahrzeugwiedererkennung.....	14
2.2.3	Ermittlung von Reisezeiten aus fahrzeuggenerierten Daten	18
2.2.4	Nachbearbeitung der Einzelfahrzeugreisezeiten	23
2.3	Empirische Analyse der verfügbaren Daten	23
2.3.1	Testfelder und Datenquellen	23
2.3.2	Aussagekraft der erhobenen Reisezeiten	25
2.3.3	Korrelationen und Dynamik verkehrlicher Kenngrößen	28
2.4	Fazit und weiteres Vorgehen	32
3	Mustererkennungsmethoden und Anwendungen zur Verkehrsprognose	33
3.1	Grundlagen und Einordnung	33
3.1.1	Stufen der Mustererkennung.....	34
3.1.2	Verfahren der Mustererkennung.....	34
3.1.3	Abgrenzung zu Verkehrsmodellen.....	35
3.2	Klassifikation und Segmentation der Datenbasis	36
3.2.1	Klassifikation nach Kalendermerkmalen.....	37
3.2.2	Segmentierung von Verkehrsmustern	37
3.3	Regressionsanalytische und zeitreihenbasierte Verfahren	38
3.4	Künstliche Neuronale Netze	40
3.5	Selbstlernende Expertensysteme	42
3.6	Instanzbasiertes Lernen	43
3.7	Optimierung der Modellparameter	47
3.8	Modellauswahl und weiteres Vorgehen	49

4	Entwicklung eines neuen Prognoseverfahrens	51
4.1	Grundlagen des Instanzbasierten Lernens.....	51
4.1.1	Die Idee.....	51
4.1.2	Anwendungsvoraussetzungen Instanzbasierter Lernverfahren.....	52
4.2	Auswahl der Merkmale und Definition der Distanzmetrik.....	53
4.2.1	Datenstruktur und Auswahl der Merkmale.....	53
4.2.2	Berechnung der Distanz D zweier Verkehrssituationen	55
4.2.3	Distanz zweier Verkehrsmuster D_V	55
4.2.4	Kalenderbasierte Distanz D_K	59
4.2.5	Strafterm S_F zur Berücksichtigung fehlender Messwerte.....	61
4.2.6	Berücksichtigung weiterer Merkmale.....	62
4.3	Anzahl Nachbarn k und lokale Approximation der Daten.....	62
4.3.1	Anzahl der zu berücksichtigenden Nachbarn k	62
4.3.2	Lokale Approximationsfunktion / Gewichtungsfunktion.....	63
4.4	Berücksichtigung aktueller Messwerte der Zielgröße	65
4.4.1	Angleichen an den aktuellen Messwert.....	66
4.4.2	Inkrementprognose	66
4.5	Prognosezuverlässigkeit	67
4.5.1	Zuverlässigkeitsindikatoren.....	68
4.5.2	Schätzung der Wahrscheinlichkeitsdichte des Prognosewertes	71
4.6	Gesamtübersicht und Implementierung	72
4.6.1	Gesamtübersicht über das Verfahren	72
4.6.2	Implementierung des Verfahrens	74
4.7	Fazit und weiteres Vorgehen.....	76
5	Parametersensitivität und Bewertung des Verfahrens	77
5.1	Grundlagen	77
5.1.1	Vorgehen.....	77
5.1.2	Methodik der Bewertung	78
5.2	Modellparametersensitivität	81
5.2.1	Analyse der einzelnen Parameter.....	81
5.2.2	Maßzahlen zur Beurteilung der Sensitivität.....	84
5.2.3	Vergleich der Sensitivitäten.....	85
5.2.4	Fazit der Sensitivitätsanalyse.....	88
5.3	Optimierung der Parameter.....	89
5.3.1	Genetischer Algorithmus zur Parameteroptimierung	89
5.3.2	Ergebnisse der Parameteroptimierung	92
5.3.3	Fazit der Parameteroptimierung.....	94
5.4	Eingesetzte Prognoseverfahren	95

5.4.1	Instanzbasierte Lernverfahren	95
5.4.2	Benchmarkverfahren	95
5.5	Ergebnisse der Belegungsgradprognose	99
5.6	Ergebnisse der Reisezeitprognose	106
5.7	Ergänzungen zur Reisezeitprognose	113
5.7.1	Test von Heuristiken zur Verbesserung der Reisezeitprognose	113
5.7.2	Anwendung der Zuverlässigkeitsindikatoren.....	115
5.7.3	Abschätzung von Vertrauensintervallen für Reisezeiten.....	116
5.8	Fazit der Bewertung	119
6	Fazit und Ausblick	121
6.1	Zusammenfassung	121
6.2	Möglichkeiten zur weiteren Entwicklung	125
	Notation und Definitionen	129
	Quellenverzeichnis.....	135
	Abbildungsverzeichnis.....	143
	Tabellenverzeichnis	147
	Anhang 1 Verfahren zur Kürzestwegsuche.....	149
	Anhang 2 Ergebnisse der Parameteroptimierung.....	151
	Anhang 3 Darstellung der Prognoseergebnisse	155

1 Einleitung

1.1 Kontext und Problemstellung

Im modernen Verkehrsmanagement des motorisierten Individualverkehrs spielen Informationen über aktuelle und zukünftige Verkehrszustände eine wichtige Rolle. Neue Technologien – insbesondere im Bereich der Telekommunikation – erlauben eine angepasste und umfassende Unterstützung des Verkehrsteilnehmers durch individuelle Dienste sowie die netzoptimale Beeinflussung der Verkehrsteilnehmer über kollektiv wirkende Systeme des Verkehrsmanagements. Dabei kommt der dynamischen Verkehrsinformation und der dynamischen Zielführung eine wichtige Bedeutung zu [BUSCH ET AL. 2006a].

Informations- und Steuerungsverfahren können ihr Potential nur dann ausschöpfen, wenn ihnen zuverlässige Informationen über den aktuellen und zukünftigen Verkehrszustand zur Verfügung stehen. Folglich bildet das Wissen über die Verkehrslage ein zentrales Element zur Sicherstellung des Nutzens und der Akzeptanz von Verkehrsmanagementmaßnahmen.

Die Verkehrslage als Abstraktum muss zur Interpretation und Weiterverarbeitung durch geeignete, durch die Anwendung und/oder die Verfügbarkeit angepasste Kenngrößen beschrieben werden. Aussagekräftige lokale Kenngrößen, wie der Belegungsgrad, können zur Charakterisierung der Verkehrssituation auf einzelnen Fahrstreifen oder Querschnitten herangezogen werden. Besondere Ausprägungen können beispielsweise zur Erkennung von Überlastungen genutzt werden. Streckenbezogene Daten stellen interessante, jedoch meist aufwändig zu erhebende verkehrliche Kenngrößen dar. Insbesondere der Reisezeit kommt eine große Bedeutung zu, da sie durch den Verkehrsteilnehmer auf seiner Fahrt erlebt wird, meist die Grundlage für Fahrt- und Routenentscheidungen darstellt und somit ein Kriterium zur Beurteilung der Qualität des Verkehrsablaufs sowohl aus Nutzer-, als auch aus System-sicht darstellt.

Für verschiedene Anwendungen und Situationen sind Verkehrskenngrößen für unterschiedliche Zeithorizonte relevant. Während für die reine Information – beispielsweise in Form einer farblich kodierten Netzdarstellung – der aktuelle Zeitpunkt relevant sein kann, gilt für Routing- und Navigationsanwendungen eine Verkehrskenngröße bereits zum Zeitpunkt der Erhebung als veraltet, da der Fahrer erst während der Fahrt, d.h. in der Zukunft, die entsprechenden Orte im Netz erreicht.

Für zukünftige Zeitpunkte liegen naturgemäß keine gemessenen Größen vor, aber auch der aktuelle Systemzustand ist in aller Regel aufgrund technischer und finanzieller Rahmenbedingungen nur unvollständig erfasst. Es werden daher Verfahren zur Schätzung und Prognose dieser Größen benötigt. Im Folgenden werden die Begriffe "Prognose" (entspricht einer Schätzung für Prognosehorizonte größer null) und "Schätzung" (entspricht einer Prog-

nose für einen Zeithorizont gleich null) synonym verwendet. Typischerweise werden in verkehrlich stark belasteten Städten und Ballungsräumen lokale Verkehrsdaten zum Zwecke der Verkehrssteuerung und des Verkehrsmanagements allgemein an Messstellen kontinuierlich erfasst und zentral zusammengeführt, verwertet und archiviert. Streckenbezogene Kenngrößen, insbesondere Reisezeiten, werden heutzutage häufig als "Abfallprodukt" aus zweckfremden Datenquellen wie Flottendispositionssystemen gewonnen und im Allgemeinen zeitlich zufällig beobachtet.

Die vorliegende Dissertation widmet sich der Frage, wie aktuelle und insbesondere zukünftige Verkehrszustände in Form von Verkehrskenngrößen geschätzt werden können. Zentral ist dabei die Nutzung von in Archiven festgehaltenen umfangreichen Datenbasen und daraus extrahierten räumlich-zeitlichen Verkehrsmustern unter Berücksichtigung der aktuellen Beobachtung.

1.2 Anforderungen an das Verfahren

Das Verfahren soll im Kontext des Verkehrsmanagements eine mögliche Grundlage für dynamische Verkehrsinformations-, Routing- und Navigationssysteme liefern. Dazu soll es verschiedene inhaltlich-funktionale und betrieblich-technische Anforderungen erfüllen.

Inhaltlich-funktionale Anforderungen

Prognostizierte Verkehrskenngrößen sollen für operatives und taktisches Echtzeit-Verkehrsmanagement auf Netzebene zur Verfügung stehen. Dazu werden Prognosen für verschiedene nutzerrelevante Horizonte benötigt. Der maximale Prognosehorizont, bei dem aktuelle Verkehrsdaten zur Prognoseberechnung berücksichtigt werden, wird zu einer Stunde angenommen.

Die räumlich-zeitlichen Verkehrszusammenhänge können sich über die Zeit ändern, z.B. wenn sich Änderungen in der Netz- und/oder der Nachfragestruktur einstellen. Diese Änderungen müssen von einem geeigneten Verfahren abgefangen werden (adaptive Komponente).

Im praktischen Einsatz können Situationen entstehen, die ein plausibles Verhalten des Prognoseverfahrens zur Laufzeit gefährden. Dazu gehören insbesondere unplausible bzw. fehlende Messwerte als Eingangsgrößen für das Verfahren, so dass die Robustheit gegenüber fehlenden Messwerten eine wichtige Eigenschaft darstellt.

Eine Prognose beinhaltet Unsicherheit. Diese liegt in der Unvorhersagbarkeit des beobachteten Prozesses, in der notwendigen Abstraktion durch ein Modell an sich und in der variablen Qualität der Eingangsgrößen des Modells begründet. Eine Information über die zu erwartende Zuverlässigkeit der Prognose kann für den Nutzer von großer Relevanz sein [BENZ ET AL. 2003]. Je nach Fahrtzweck kann die geschätzte Prognosezuverlässigkeit bei der individuellen

Planung bzw. durch entsprechende Routingverfahren berücksichtigt werden [KAPARIAS ET AL. 2007]. Eine Anforderung an das Verfahren ist die Möglichkeit, Indikatoren zur Abschätzung der Prognosezuverlässigkeit aus modellinhärenten Größen ableiten zu können.

Betrieblich-technische Anforderungen

Die Einsetzbarkeit im praktischen Betrieb stellt eine wesentliche Grundanforderung an das Prognoseverfahren dar und muss schon bei der Konzeption bedacht werden. Der Erfolg eines Prognoseverfahrens hängt unter anderem und wesentlich von der Verfügbarkeit der benötigten Daten ab. Sind Eingangsdaten praktisch nicht verfügbar (z.B. aufgrund nicht marktreifer Technologien, hoher Kosten oder unüberwindbarer administrativer Hürden), so ist der Einsatz des Prognoseverfahrens über eine Demonstration hinaus unwahrscheinlich. Um praktisch relevant zu sein, soll das Verfahren auf nach dem heutigen Stand der Technik verfügbaren Daten basieren.

Der Aufwand für den zuverlässigen Betrieb des Verfahrens sollte nicht zu groß sein. Dabei lässt vor allem eine aufwändige Modellkalibrierung ein theoretisch gut funktionierendes Verfahren in der Praxis aufgrund fehlender Ressourcen oder fehlender Daten zur Kalibrierung häufig scheitern. Der Parametrieraufwand sollte daher möglichst gering sein.

Schließlich sollte ein Prognoseverfahren möglichst übertrag- und erweiterbar sein. Mittelfristig können sich neben dem "Entwicklungs- und Testfeld" weitere Einsatzräume ergeben, wodurch die Portierbarkeit auf andere Netze erforderlich wird. Auch kann die technologische Entwicklung und sich ändernde technische oder administrative Randbedingungen die Anpassung der bestehenden Datenerfassung oder die Ergänzung neuer Datenquellen zur Folge haben. Das Verfahren soll somit auf andere Netze und Randbedingungen möglichst einfach übertragbar sein.

1.3 Lösungsansatz und methodischer Aufbau

Zur Lösung des Prognoseproblems wird ein Instanzbasiertes Lernverfahren entwickelt. Dem Instanzbasierten Lernen liegt, wie bei anderen Mustererkennungsverfahren auch, die Annahme zugrunde, dass sich die aktuelle Entwicklung einer Verkehrskenngröße aus der Entwicklung einer der aktuellen Situation ähnlichen, im Datenarchiv vorliegenden Situation ("Instanz") ableiten lässt.

Im Sinne des Regressionsgedankens wird die interessierende Verkehrskenngröße als zufällig beobachtete abhängige Variable durch kontinuierlich erfasste unabhängige Variable prognostiziert. Diese Formulierung erlaubt die Prognose von zufällig beobachteten Größen (z.B. Reisezeiten aus fahrzeuggenerierten Daten) und von kontinuierlich erfassten Daten (z.B. lokale Belegungsgrade) mit demselben Mechanismus. Die unabhängigen Variablen

werden in Merkmalsvektoren zusammengefasst und müssen für eine kontinuierliche Prognose zum einen zeitlich vollständig vorliegen und zum anderen die Verkehrssituation bezüglich ihrer nahzukünftigen Entwicklung gut beschreiben, also gute Prädiktoren sein. Als unabhängige Variable werden in dieser Arbeit aktuelle und "nahvergangene" Verkehrsdaten lokaler Messstellen verwendet, da diese in relevanten Netzen typischerweise kontinuierlich erfasst werden und einen Rückschluss auf die lokale Verkehrssituation zulassen. Als weitere unabhängige Variable dienen Kalendermerkmale (Wochentag, Uhrzeit, Veranstaltungen), da diese insbesondere für größere Horizonte als aussagekräftige Prädiktoren genutzt werden können. Das Verfahren erlaubt die Integration weiterer beschreibender Kenngrößen.

Die zu prognostizierenden abhängigen Variablen sind durch Ort und Zeit (Prognosehorizont) definiert. Ein wesentlicher Aspekt ist die Gewichtung der einzelnen Merkmale bei der Berechnung der Ähnlichkeit des aktuellen Musters mit den archivierten Instanzen. Die Gewichtung hängt sowohl von der räumlichen als auch von der zeitlichen Lage des Merkmals zur abhängigen Variable ab und wird auf Basis von geometrischen Kenngrößen im Netz und dem Erfassungszeitpunkt funktional ermittelt.

Arbeitsschritte und methodischer Aufbau der Arbeit, sowie der Ablauf des Verfahrens sind in Abb. 1.1 dargestellt.

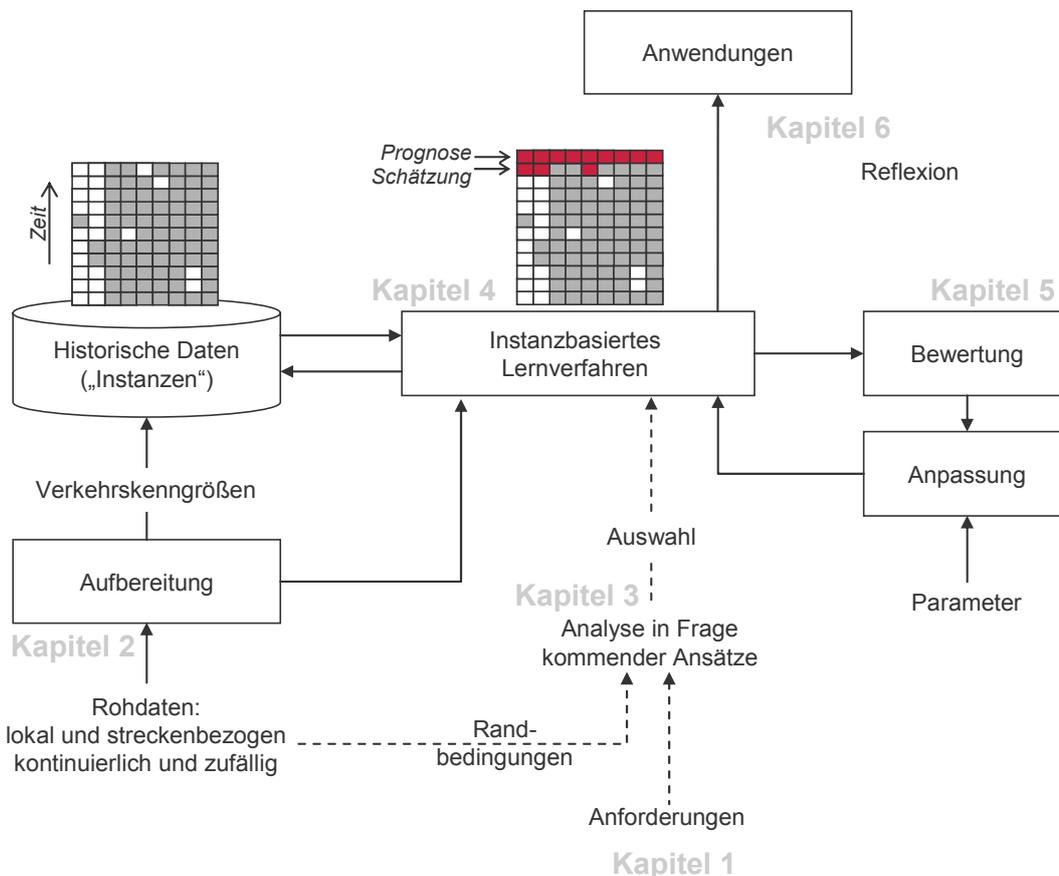


Abb. 1.1 Methodischer Aufbau der Arbeit und des Verfahrens

In *Kapitel 1* sind die Anforderungen an das Prognoseverfahren aus funktionaler, technischer und betrieblicher Sicht definiert worden.

In *Kapitel 2* werden Technologien für die lokale und streckenbezogene Verkehrsdatenerfassung beschrieben und Methoden zur Aufbereitung der Daten zu verkehrlichen Kenngrößen dargestellt. Dabei kommt der Aufbereitung und Filterung von streckenbezogenen Daten (Fahrzeugpositionsmeldungen und Fahrzeugwiedererkennung) zu Reisezeiten eine große Bedeutung zu. Abhängig von ihrer typischen zeitlichen Verfügbarkeit können die Daten als unabhängige Variable (kontinuierlich verfügbar) oder abhängige Variable (mindestens zufällig beobachtet) in dem zu entwickelnden Prognoseverfahren verwendet werden. Abschließend werden einige empirische Untersuchungen zu räumlichen und zeitlichen Zusammenhängen und Eigenschaften der Daten und Kenngrößen auf Basis der in den Testfeldern Graz und München erhobenen Messwerte durchgeführt.

Auf Basis der definierten Anforderungen und der Eigenschaften der in der Praxis zur Verfügung stehenden Datenquellen werden in *Kapitel 3* die prinzipiell in Frage kommenden Klassen von Prognoseverfahren analysiert. Unter diesen, allgemein aus dem Bereich der "Datenmustererkennung" stammenden Verfahren, wird mit dem Instanzbasierten Lernen ein geeignetes ausgewählt. Wesentlich für die Auswahl sind dabei die Transparenz und die zu erwartende Robustheit des Verfahrens, sowie vor allem die Möglichkeit, relevante statistische Kenngrößen zur Prognosezuverlässigkeit direkt aus den Daten abzuleiten.

In *Kapitel 4* wird das Verfahren bezüglich seiner Eingangsgrößen, Verarbeitungsschritte und Parameter spezifiziert. Im Prinzip werden mehrdimensionale Merkmalsvektoren aus verkehrlichen und externen Variablen definiert, anhand derer die Ähnlichkeit von zwei (Verkehrs-) Situationen bewertet werden kann. Anschließend werden Möglichkeiten zur Quali- und Quantifizierung der Prognosezuverlässigkeit dargestellt und die Implementierung des Verfahrens beschrieben.

In *Kapitel 5* wird zunächst die Sensitivität der Verfahrensergebnisse bezüglich der sechs Verfahrensparameter untersucht. Aufgrund zu erwartender Wechselwirkungen wird statt einer manuellen Kalibrierung der Parameter mit einem Genetischen Algorithmus ein automatisches Verfahren zur Parametereinstellung vorgeschlagen und exemplarisch angewendet. Anschließend wird das Verfahren für verschiedene Szenarien bewertet und mit anderen Verfahrenstypen verglichen. Schließlich werden die in Kapitel 4 beschriebenen Möglichkeiten zur Schätzung der Prognosezuverlässigkeit anhand verschiedener Situationen untersucht.

In *Kapitel 6* werden die Erkenntnisse der Arbeit zusammengefasst und reflektiert, verschiedene Anwendungsfelder aufgezeigt und mögliche Entwicklungsrichtungen beschrieben.

2 Verkehrsdatenerfassung und -aufbereitung

In *Kapitel 2* werden Technologien für die lokale und streckenbezogene Verkehrsdatenerfassung beschrieben und Methoden zur Aufbereitung der Daten zu verkehrlichen Kenngrößen dargestellt. Dabei kommt der Aufbereitung und Filterung von streckenbezogenen Daten (Fahrzeugpositionsmeldungen und Fahrzeugwiedererkennung) zu Reisezeiten eine große Bedeutung zu. Abhängig von ihrer typischen zeitlichen Verfügbarkeit können die Daten als unabhängige Variable (kontinuierlich verfügbar) oder abhängige Variable (mindestens zufällig beobachtet) in dem zu entwickelnden Prognoseverfahren verwendet werden. Abschließend werden einige empirische Untersuchungen zu räumlichen und zeitlichen Zusammenhängen und Eigenschaften der Daten und Kenngrößen auf Basis der in den Testfeldern Graz und München erhobenen Messwerte durchgeführt.

2.1 Erfassung und Aufbereitung lokaler Verkehrsdaten

2.1.1 Lokale Verkehrsdatenerfassung

Lokale Verkehrsdaten beschreiben den Verkehrszustand an einer Messstelle oder einem Messquerschnitt. Wichtige gemessene Kenngrößen sind dabei die Verkehrsstärke q [Fz/h], der Belegungsgrad b [%] und die Geschwindigkeit v [km/h]. Diese werden vor Ort im Steuergerät oder in der Streckenstation aus den Sensorsignalen einzelner Fahrzeuge ermittelt und meist auf fixe Zeitintervalle (z.B. 60 oder 90 Sekunden) bezogen.

Je nach Detektortyp werden die Größen direkt aus den physikalischen Signalen der Sensorik abgeleitet oder unter Zuhilfenahme zusätzlicher Annahmen ermittelt. Letzteres trifft insbesondere innerorts für die Geschwindigkeitsermittlung mit den häufig eingesetzten Einzel-Induktivschleifendetektoren zu. Während q und b direkt aus der Verstimmung des erzeugten elektrischen Feldes abgeleitet werden, wird v über die Annahme einer Standardfahrzeuglänge aus q und b ermittelt.

Aus q und v kann die lokale Dichte $\rho = q/v$ berechnet werden. Diese ist über einen konstanten Faktor, der abhängig von der Fahrzeuglänge ist, wiederum mit b verknüpft. Darüber hinaus können weitere Kenngrößen zur lokalen Beschreibung der Verkehrslage ermittelt werden (beschrieben beispielsweise in den TLS 2006 für den Außerortsbereich).

Zu nennen sind hier vor allem die Standardabweichung der Einzelgeschwindigkeiten s [km/h], die mittlere Nettozeitlücke t_{netto} [s] und die zukünftig (eventuell) verfügbare Time To Collision TTC [s]. Sämtliche dieser Größen beschreiben das Abstands- und Geschwindigkeitsverhältnis der einzelnen Fahrzeuge zueinander und können über q , v und b hinaus

wertvolle Hinweise zur Verkehrssituation geben (z.B. sich anbahnende Störungen durch erhöhte Variabilität der Geschwindigkeiten).

Zur Erfassung lokaler Verkehrsdaten werden verschiedene Technologien eingesetzt. Innerorts werden bzw. wurden Detektoren hauptsächlich zum Zwecke der lokalen verkehrabhängigen Lichtsignalanlagensteuerung verwendet. Als so genannte "LSA Detektoren" sind heute i.d.R. Einzel-Induktivschleifendetektoren eingesetzt. Außerorts sind im Bereich von Verkehrsleiteinrichtungen und als Dauerzählstellen zu statistischen Erhebungen meist Doppel-Induktivschleifen im Einsatz. Bei Doppel-Induktivschleifen kann über den Zeitversatz zwischen den Überfahrten der beiden Schleifen und dem bekannten Abstand die Geschwindigkeit direkt ermittelt werden. Doppel-Induktivschleifen werden auch innerorts an strategisch wichtigen Querschnitten eingesetzt (sogenannte "Strategische Detektoren"). Sowohl innerorts als auch außerorts kommen zunehmend Überkopfdetektoren zum Einsatz. Diese direkt über der Fahrbahn oder von schräg oben messenden Detektoren arbeiten meist mit Ultraschall-, Infrarot- und/oder Radarsensoren. Auch Aktiv-Laser Scanner werden eingesetzt. Innerorts und in beleuchteten Bereichen (z.B. Tunneln) kommen außerdem Videodetektoren zum Einsatz (siehe MESCHENDÖRFER [2004] für einen Überblick über die unterschiedlichen Detektorarten).

2.1.2 Beschreibung des lokalen Verkehrszustands

Lokal erfasste Verkehrskenngrößen sollten tatsächliche Verkehrssituationen möglichst gut und eindeutig beschreiben. Das gilt sowohl für die Nutzung als unabhängige Variable als auch als Prädiktorvariable.

Dazu sollten nach Möglichkeit direkt gemessene und nicht unter zusätzlichen Annahmen ermittelte Größen verwendet werden. Da das beschriebene Prognoseverfahren auch und vor allem innerorts eingesetzt werden soll und die zur Entwicklung verwendete Datengrundlage hauptsächlich mittels Einzel-Induktivschleifen erfasst wurde, wird im Folgenden von den gemessenen Verkehrsstärken q und Belegungsgraden b ausgegangen. Aufgrund von Effizienz und Speicherplatzbedarf soll dabei möglichst wenig redundante Information weiterverarbeitet werden. Ziel ist es, *eine* Verkehrskenngröße zu identifizieren, die den lokalen Verkehrszustand möglichst gut beschreibt.

Belegung und Verkehrsstärke sind im frei fließenden Verkehr näherungsweise linear voneinander abhängig: Je höher die Verkehrsstärke, desto höher die Belegung. Streuungen ergeben sich durch unterschiedliche Fahrzeuglängen. So führen Lkw aufgrund der größeren Länge bei gleicher Geschwindigkeit und Verkehrsstärke zu einer stärkeren Belegung als Pkw.

2.1.3 Aufbereitung lokaler Verkehrsdaten

Die lokal gemessenen Verkehrsgrößen werden vor der Weiterverarbeitung aufbereitet. Ziel ist es, nicht verwendbare Messwerte zu erkennen und herauszufiltern, sowie stochastische und somit verkehrlich unbedeutende Schwankungen in der Zeitreihe zu eliminieren und wesentliche verkehrliche Phänomene (z.B. Stauaufbau und -abbau) durch eine geeignete Aggregation/Glättung zu erhalten bzw. hervorzuheben.

Unplausible/nicht verwendbare Messwerte

Von der Erfassung der Fahrzeuge auf der Straße bis zur Bereitstellung der Messwerte in der Zentrale können verschiedene technische Fehlfunktionen zu fehlerhaften oder fehlenden Daten führen. Abb. 2.1 zeigt Belegungsgrade für im Netz verteilte Detektoren über die Zeit. In Weiß dargestellt sind fehlende Messwerte; ein Detektor liefert offensichtlich unplausible Werte.

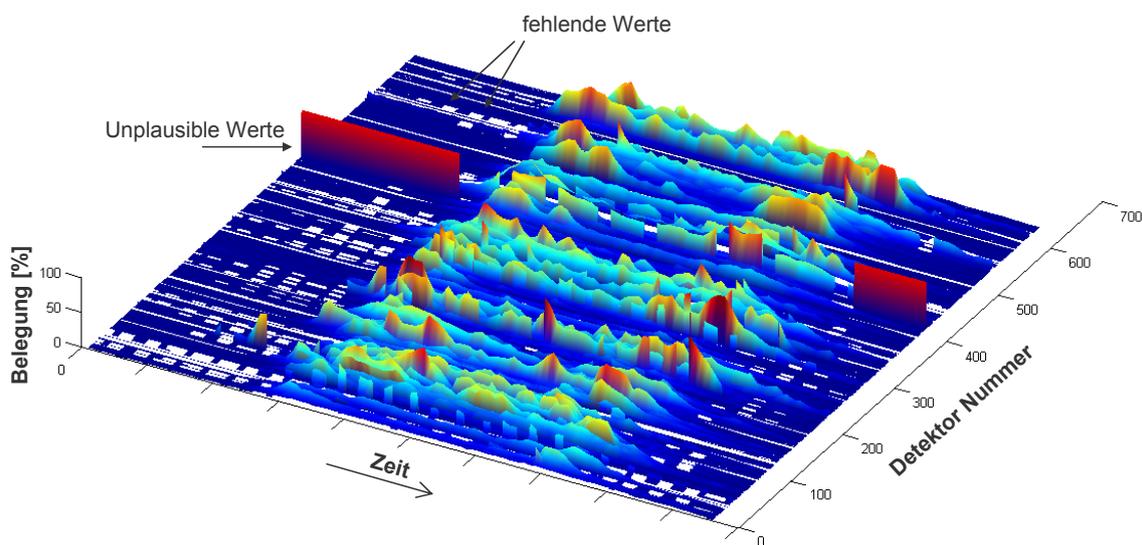


Abb. 2.1 Darstellung des Belegungsgrades (Median über 15 Minuten) je Detektor über die Zeit.
Graz, Mittwoch 02.05.2007

Die entsprechenden Zeitintervalle müssen bei der Datenvorverarbeitung möglichst vollständig erkannt und für die weitere Verarbeitung (Aggregation, Glättung) entsprechend markiert werden. Methodisch einfach ist dabei die Erkennung fehlender Daten und von Daten, die bereits in vorgeschalteten Systemen mittels Fehlerkennung als fehlerhaft markiert wurden.

Zur Identifizierung unplausibler Daten wurden Verfahren zur automatisierten Plausibilisierung lokaler Messgrößen entwickelt. Verfahren für den Einsatz im Bereich von Schnellstraßen sind dabei in der Überzahl (siehe z.B. BUSCH ET AL. 2006b, HOOPS 2003). Zum einen bieten streckenbezogene Kriterien, wie z.B. die Fahrzeugerhaltung zwischen zwei Messquerschnitten ohne Ein- und Ausfahrten, die Möglichkeit, über die Bilanzierung potentielle Messfehler

herauszufinden. Dies ist in einem vermaschten innerstädtischen Straßennetz i.d.R. nicht möglich. Zum anderen sind auf Schnellstraßen fundamentale Zusammenhänge der makroskopischen Verkehrskenngrößen deutlich charakteristischer als in städtischen Netzen, wo diese aufgrund verschiedener Störeinflüsse (insbesondere Knotenpunkte) nicht oder nicht sehr ausgeprägt beobachtbar sind.

Aus diesen Gründen ist man in städtischen Straßennetzen bezüglich der Möglichkeiten eingeschränkt. Eingesetzte Verfahren beruhen häufig auf der Auswertung von Schwellenwerten, regressionsbasierten Ausreißertests zum Feststellen von Abweichungen im Fundamentaldiagramm (meist handelt es sich dabei um ein q - b Diagramm) und Abweichungen von typischen Ganglinien (siehe beispielsweise FREUDENBERGER 2001). Abweichungstests für Fundamentaldiagramme und Ganglinien sind sehr leistungsfähige, jedoch auch sensible Werkzeuge. Ihr Einsatz und ihre Anpassung an lokale Gegebenheiten muss im Prinzip für jede Messstelle gesondert durchgeführt und geprüft werden. Da dies den Rahmen dieser Arbeit sprengen würde, werden nur zwei einfache Filter verwendet:

1. *Hängenbleiben*: Wenn über mehrere Zeitintervalle (hier: 5 Intervalle) ein konstanter, also identischer, Wert geliefert wird, dann sind die Daten aller Wahrscheinlichkeit nach unplausibel.
2. *System-Fehlercode*: Das System liefert einen Fehlercode (hier: b : 255/ q : 65535) und die Daten sind somit offensichtlich nicht nutzbar. Wenn eine korrekte Ermittlung des Messwertes nicht möglich ist, wird ein Fehlercode erzeugt und der Wert nicht für die weitere Verarbeitung genutzt.

Unplausible Daten können trotz Prüfung und auch bei Anwendung komplexerer Filter unerkannt bleiben. Bei der weiteren Verarbeitung sorgen folgende Mechanismen und Eigenschaften dafür, dass der Einfluss fehlerhafter Messungen gering gehalten wird:

- Eine Aggregation der Daten auf 15 Minuten und eine anschließende exponentielle Glättung sorgen dafür, dass einzelne fehlerhafte Messungen ("zufällige Fehler") nicht stark ins Gewicht fallen.
- Systematische Fehler (z.B. Belegung immer um 10% zu hoch) der unabhängigen Variablen stellen in einem Mustererkennungsverfahren kein Problem dar.
- Fehlerhafte Daten einer Messstelle können aufgrund der Verwendung der Daten mehrerer Messstellen als unabhängige Variable bis zu einem gewissen Grad ausgeglichen werden.

Ein Problem könnten dynamische systematische Fehler (also z.B. ein stetiges Anwachsen des Messfehlers) und an einer Stelle häufig auftretende zufällige Fehler (die dann im eigentlichen Sinn nicht mehr zufällig wären) sein. Hierfür wäre der Einsatz robuster Erkennungs-

methoden sinnvoll und ein aufmerksamer Operator zur "manuellen Erkennung" wünschenswert.

Aggregation und Glättung lokaler Verkehrsdaten

Eine Herausfilterung stochastischer Schwankungen kann durch eine geeignete Aggregation und/oder Glättung der Zeitreihe erreicht werden. In der Literatur werden verschiedene Verfahren zur Glättung von Verkehrsdaten vorgeschlagen (siehe z.B. [BREITENSTEIN 1978], [MARZ 1999]).

In verkehrstechnischen Anwendungen werden häufig die Aggregation bzw. Mittelwertbildung bezogen auf ein größeres Zeitintervall eingesetzt. Dazu wird ein fixes Zeitraster der Breite Δt vorgegeben. Zweckmäßigerweise entspricht dies dem später zur Ablage der Verkehrsmuster verwendeten Raster. Der *aggregierte Wert* für das Intervall $]t, t + \Delta t]$ ermittelt sich zu:

$$x_{agg}(t, t + \Delta t) = \int_{\tau=t}^{t+\Delta t} x(\tau) d\tau$$

Handelt es sich um eine nicht abzählbare Größe (beispielsweise b), so muss der Wert noch durch die Anzahl der verwendeten Einzelwerte geteilt werden (= Mittelwert).

Durch die Aggregation auf fixe Zeitintervalle werden die Daten in ein einheitliches Zeitraster gebracht, die Datenmenge reduziert und eine Glättung erreicht. Die nun vorliegenden aggregierten Daten können einer weiteren Glättung unterworfen werden. Ein häufig eingesetztes Verfahren ist die gleitende Mittelwertbildung über die letzten n Intervalle und die exponentielle Glättung als deren Sonderform. Der *einfache gleitende Mittelwert* berechnet sich zu:

$$x_{glatt}(t) = \frac{1}{n} \cdot \sum_{i=0}^{n-1} x(t-i)$$

Die Gleichung kann noch um einen Gewichtungsfaktor $g(i)$ zur Berücksichtigung der zeitlichen Distanz zum Messwert zum *gewichteten gleitenden Mittelwert* erweitert werden.

$$x_{glatt,gewichtet}(t) = \frac{1}{n} \cdot \sum_{i=0}^{n-1} g(i) \cdot x(t-i)$$

Wird $g(i)$ mit steigendem i monoton fallend gewählt wird eine stärkere Gewichtung der aktuellen Messwerte erreicht. Ein Sonderfall des gewichteten gleitenden Mittelwertes ist das Verfahren der exponentiellen Glättung (oder auch: *exponentiell gewichtete gleitende Mittelwerte*). Der Vorteil ist hier, dass sich die Reihe rekursiv aufbaut und nur der letzte geglättete Wert gespeichert werden muss.

$$x_{glatt,exp}(t) = \alpha \cdot x(t) + (1 - \alpha) \cdot x_{glatt,exp}(t - 1)$$

mit $0 \leq \alpha \leq 1$ und $x_{glatt,exp}(0) = x(0)$ als Startwert

Der exponentiell geglättete Wert enthält damit Informationen aller vorhergehenden Zeitintervalle, die jedoch jeweils nur mit dem Gewicht $(1 - \alpha)^t$ eingehen. Je größer der Glättungsfaktor α gewählt wird, desto stärker ist die Glättung.

Für Aggregation und Glättung müssen Parameter gewählt werden. Für die optimale Wahl der Glättungsparameter gibt es keine Regel. In Abhängigkeit von der Anwendung müssen sie vom Nutzer plausibel eingestellt werden. Ziel der Aggregation und Glättung ist es, die stochastischen Schwankungen herauszufiltern und gleichzeitig verkehrlich relevante Phänomene in der Zeitreihe zu erhalten. Als nicht ungünstig für die weitere Verarbeitung hat sich eine Aggregation auf 15 Minuten und eine anschließende exponentielle Glättung mit $\alpha = 0,3$ herausgestellt (ein Vergleich verschiedener α ist in Abb. 2.2 dargestellt).

Neben der zeitlichen Vorverarbeitung könnte auch eine räumliche Aggregation durch das Zusammenfassen mehrerer an einem Querschnitt liegenden Messstellen (Detektoren) zu einem Querschnittswert durchgeführt werden. Da so jedoch Information verloren gehen könnte (z.B. eine außergewöhnlich hohe oder niedrige Belastung einer der Fahrstreifen) wird darauf verzichtet. Die Ergebnisse sind also fahrestreifenbezogene zeitlich aggregierte und geglättete Messwerte.

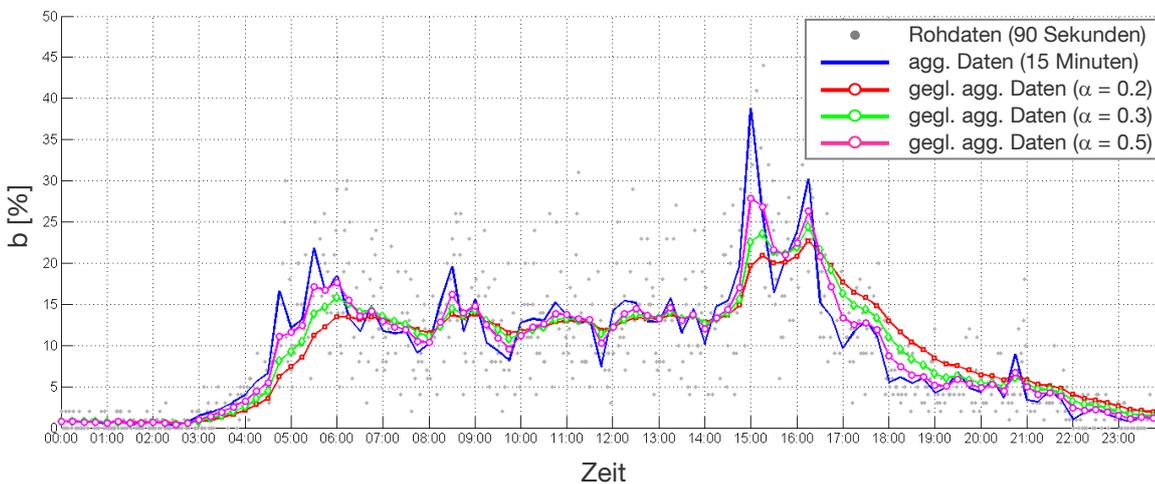


Abb. 2.2 Lokaler Belegungsgrad über Zeit (21.06.2007. Det. ID 622). Darstellung von Rohdaten, aggregierten Daten und exponentiell geglätteten Daten für verschiedene α

2.2 Erfassung und Aufbereitung der Reisezeit

Aufgrund der einfachen Interpretierbarkeit stellt die Reisezeit eine wichtige Eingangsgröße für verschiedene verkehrstelematische Anwendungen wie Informations- und Navigations-

systeme dar. Die Reisezeit spielt insbesondere in mittleren und großen Städten und in Ballungsräumen mit mehr oder minder stark vermaschten Straßennetzen und Alternativrouten als Entscheidungsgrundlage eine Rolle.

Die "klassische Art" der dynamischen Reisezeitermittlung ist der Einsatz von Verkehrsmodellen (oder besser *Verkehrsablaufmodellen*), welche die Reisezeit auf Basis physikalischer Gesetzmäßigkeiten aus lokal erfassten Verkehrsdaten und Annahmen über die Verkehrsnachfrage und die Verteilung der Fahrzeuge im Netz schätzen. Obwohl methodisch ansprechend, stellt die mangelnde flächendeckende Verfügbarkeit, Aktualität und Qualität der benötigten Eingangsdaten insbesondere innerorts häufig ein Hindernis für einen erfolgreichen praktischen Einsatz dar. So stellt z.B. ROBINSON [2005, S. 126] auf Basis einer vertieften Analyse fest:

"The biggest criticism of the queuing based models is that they rely either on a large number of data inputs, or on data inputs which can be difficult to measure accurately. As alluded to above, saturation flow and capacity flow can be difficult to measure directly. Even the traffic signal settings such as cycle time, stage times, and offsets may be difficult to collect."

Nichtsdestoweniger können modellierte Reisezeiten natürlich auch in Verfahren der Mustererkennung zur Prognose weiter genutzt werden. Jedoch würde selbst eine sehr oberflächliche Betrachtung aufgrund der Komplexität und Vielzahl an Modellen den Rahmen dieser Arbeit sprengen. Es existiert eine Vielzahl von Veröffentlichungen, z.B. HOOGENDOORN UND BOVY 2001, HELBING 1997, DAGANZO 1997, KERNER 2004 (Fokus außerorts) und BELL ET AL. 1997, FELLENDORF ET AL. 2000, LOGI UND ULRICH 2001 (Fokus innerorts).

Aktuelle und zukünftige Erfassungs- und Kommunikationstechnologien erlauben eine Ermittlung der Reisezeit direkt aus räumlich-zeitlich verteilten Fahrzeugdaten (mobil und stationär). Im Folgenden werden entsprechende Technologien und Methoden und deren Aufbereitung vertieft betrachtet.

2.2.1 Grundlagen

Die tatsächliche und messbare Reisezeit ist im Gegensatz zur momentanen Geschwindigkeit oder der daraus ableitbaren momentanen Reisezeit (*engl.: instantaneous traveltime*) eine tatsächlich "erfahrbare" Größe und beinhaltet die zeitliche Komponente einer Fahrt. Sie ist bezogen auf eine Route (meist bestehend aus einer oder mehreren Kanten) im betrachteten Straßennetz und kann zeitlich der Start- oder Ankunftszeit zugeordnet werden. Für die Anwendung im Verkehrsmanagement ist die Reisezeit zum Startzeitpunkt relevant ("wenn in die Route um 08:00 Uhr eingefahren wird, dann beträgt die Reisezeit 10 Minuten").

Reisezeiten variieren von Verkehrsteilnehmer zu Verkehrsteilnehmer, insbesondere bei frei bestimmbarer Geschwindigkeitswahl, z.B. auf Autobahnen, und in städtischen Straßennetzen durch Warteprozesse an Knotenpunkten ("*grün gerade noch erwischt?*").

Durch die gängige Technologie erfassbar ist die individuelle Reisezeit TT eines Fahrzeugs bzw. einer Fahrt i auf Route r :

$$TT_r^i = T_{r,Ankunft}^i - T_{r,Start}^i$$

Bezogen auf ein Zeitintervall kann die mittlere realisierte Reisezeit TT auf einer Route r bezogen auf das Startzeitintervall $]t, t+\Delta t]$ ermittelt werden:

$$TT_r^{]t, \Delta t]} = \frac{1}{n} \sum_{i=1}^n TT_r^i \quad \forall i \quad \left| \quad t < T_{r,Start}^i \leq t + \Delta t$$

Die mittlere realisierte Reisezeit lässt sich bei Vollerfassung ermitteln und diese wird – bei einer ausreichenden Anzahl an Fahrzeugen – eine gute Schätzung für die zu erwartende Reisezeit abgeben.

Die erhobenen Reisezeiten stellen als realisierte Reisezeit jeweils eine obere Schranke, jedoch keine untere Schranke für die mindestens benötigte Reisezeit dar. Eine hohe Reisezeit lässt also nicht zwingend auf einen Stau schließen, da sie beispielsweise durch einen freiwilligen Halt am Straßenrand zustande kam. Andersherum lässt eine niedrige Reisezeit den Schluss zu, dass günstige Verkehrsbedingungen geherrscht haben. Eine Ausnahme bildet legal oder illegal von der Straßenverkehrsordnung abweichendes Fahrverhalten (beispielsweise durch ein Polizeifahrzeug). Dies ist ggf. durch entsprechende Filter zu erkennen und auszuschließen. In der Praxis werden diese Situationen aufgrund des seltenen Auftretens keine große Rolle spielen.

Bei einer großen Anzahl erfasster Fahrzeuge lassen sich, wie an späterer Stelle beschrieben, durch Filter und Aggregierungsmethoden untypische Fahrten herausfiltern. Einzelne Beobachtungen lassen sich nur schwer bzw. nicht plausibilisieren, und müssen als das verwendet werden, was sie sind: Einzelbeobachtungen. Das in dieser Arbeit entwickelte Instanzbasierte Lernverfahren hat als Regressionsmethode die günstige Eigenschaft, veräuschte Daten zu glätten (näheres siehe Kapitel 4).

2.2.2 Erhebung von Reisezeiten mittels Fahrzeugwiedererkennung

Fahrzeugwiedererkennung bietet die Möglichkeit, die Reisezeit zwischen zwei Erfassungspunkten in einem Straßennetz direkt über die Verweilzeit der Fahrzeuge innerhalb des Abschnittes zu messen. Der wesentliche Verfahrensschritt ist dabei die Erkennung von Fahr-

zeugmerkmalen, die eine möglichst eindeutige Wiedererkennung erlauben. Der prinzipielle Aufbau einer Anordnung zur Fahrzeugwiedererkennung ist in Abb. 2.3 dargestellt.

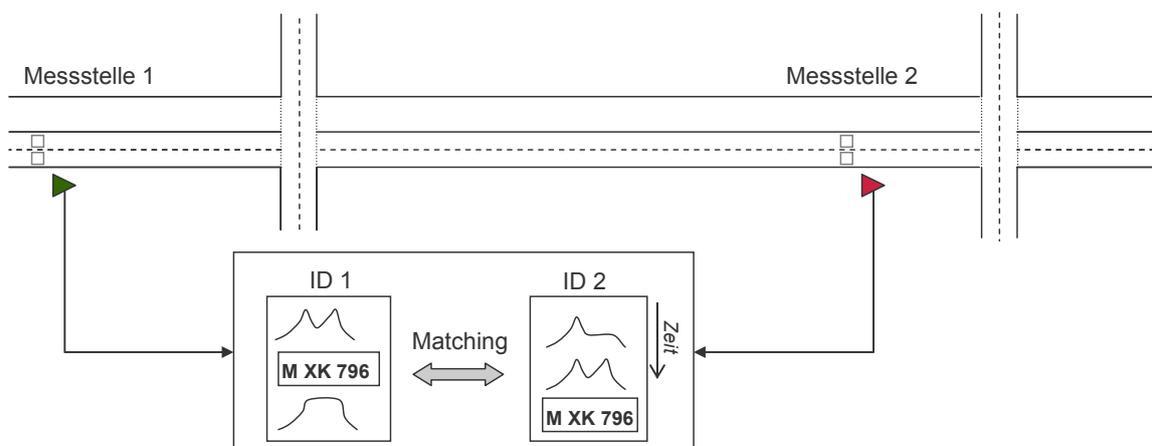


Abb. 2.3 Fahrzeugwiedererkennung (prinzipielle Systemanordnung)

Fahrzeugwiedererkennung über Fahrzeug-Infrastruktur Kommunikation

Fahrzeug-Infrastruktur (engl.: *Vehicle to Infrastructure, V2I*) Kommunikation zeichnet sich durch die Interaktion von Sende- und Empfangseinrichtungen in Fahrzeugen und entlang der Straßen aus. In den 1980er und 1990er Jahren wurde das Projekt und das System "LISB" (Leit- und Informationssystem Berlin) in Berlin entwickelt und demonstriert. Damals wurden etwa 700 Testfahrzeuge mit "Ali-Scout" On-board Units ausgestattet und etwa 250 Road Site Units (RSUs) im Berliner Straßennetz installiert. Damit waren etwa 3000 km des Straßennetzes abgedeckt (LERNER 2005).

Als Technologien kommen zur V2I Kommunikation unter Anderem Radio Frequency Identification (RFID) Tags und Wireless Local Area Network (WLAN) in Frage. Als sehr interessant kann die infrastrukturseitige Identifizierung von in Fahrzeugen mitgeführten und sichtbaren *Bluetooth* Schnittstellen angesehen werden. Sobald Bluetooth-Geräte in Betrieb gesetzt werden, identifizieren sie sich über eine weltweit einzigartige Seriennummer, so dass eine eindeutige Identifikation möglich ist. Im Standby-Modus "lauschen" unverbundene Geräte in Abständen von bis zu 2,56 Sekunden nach Nachrichten (sog. Scan Modus). Dazu müssen die Geräte aktiv auf *sichtbar* ("enabled") gestellt sein. Die Reichweite von Bluetooth beträgt üblicherweise 10-100 Meter. Im Rahmen des Projektes BLIDS [WEINZERL 2006] wurde ein Verfahren zur Verkehrserfassung entwickelt, das auf im Verkehrsfluss befindlichen Geräten mit aktivierter Bluetooth Schnittstelle basiert. Um die Praxistauglichkeit zu überprüfen, wurden an einer Straßenkreuzung in der Stadt Graz vier Bluetooth Access Points installiert. Dazu wurden über einen Zeitraum von drei Monaten Fahrzeugverfolgungen basierend auf der Bluetooth-Adresse durchgeführt, wobei ein Anteil der erfassten Geräte mit aktivierter Bluetooth - Schnittstelle von etwa 5% beobachtet wurde. Der Vorteil der Fahrzeugwiedererkennung über Bluetooth ist, dass ausschließlich infrastrukturseitige Investitionen anfallen. Das

Verfahren erlaubt eine anonyme Wiedererkennung von Verkehrsteilnehmern, da an Hand der Bluetooth-Daten kein Rückschluss auf den Besitzer des Gerätes möglich ist. Im Falle eines Mobiltelefons sind die persönlichen Daten auf der SIM-Karte gespeichert, die weder an ein bestimmtes Gerät gebunden ist noch einen Bezug zur Bluetooth-Adresse hat.

V2I Kommunikationssysteme befinden sich derzeit in den meisten Fällen in Entwicklungs- und Demonstrationsphasen. Im Zuge neuer sicherheits-, effizienz- und umweltrelevanter fahrzeugseitiger Steuerungs- und Informationsapplikationen wird die benötigte Hardware vermutlich zukünftig verfügbar werden, so dass mittels V2I Systemen erfasste Daten als relevant betrachtet werden können.

Wiedererkennung über Detektorsignale ("Signaturen")

Eine weitere Möglichkeit zur Fahrzeugwiedererkennung ist die Nutzung von Signalen aus für die lokale Verkehrsdatenerfassung eingesetzten Detektoren als Muster zur Zuordnung eines Fahrzeuges an mehreren Messquerschnitten. Obgleich sich jede Sensortechnologie, die ein entsprechend detailliertes Signal erfasst, dazu eignet, werden heute hauptsächlich Signale von Induktivschleifendetektoren genutzt.

BÖHNKE [1980] und PFANNERSTILL [1983] haben sich bereits in den frühen 1980er Jahren dem Thema gewidmet und ein Verfahren zur Wiedererkennung von Fahrzeuggruppen auf Autobahnen entwickelt. Die Forschungen sind in das Produkt MAVE eingeflossen, das derzeit einzige Produkt, das kommerziell vertrieben und eingesetzt wird (z.B. Verkehrsbeeinflussungsanlage A8 Hohenstadt-Riedheim, [AVE 2007]).

RITCHIE UND SUN [2001] und MAIER ET AL. [2007] stellten Verfahren zur Wiedererkennung von Einzelfahrzeugen im innerörtlichen Kontext vor. Dabei werden die Signaturen nach einem Normierungs- und Filterungsprozess jeweils bezüglich der Ähnlichkeit der Signaturen und deren Ableitungen verglichen. Das Verfahren von MAIER ET AL. [2007] wurde im innerörtlichen Kontext unter erschwerten Bedingungen (Zu- und Abflussrate über 85%) getestet. Dabei konnte etwa jedes achte Fahrzeug richtig wiedererkannt werden.

COIFMAN [1998, 2001, 2002] entwickelte ein Verfahren zur Wiedererkennung von Fahrzeugen, wobei als Fahrzeugmerkmal allein die magnetische Länge genutzt wird. Zusätzlich wird die Information über die Position der Fahrzeuge in einer so genannten "Vehicle Match Matrix" genutzt. STEINER UND SICK [2007, 2008] entwickelten und testeten einen Ansatz speziell für den Fall unbekannter Zu- und Abflüsse sowie größerer Durchmischungsraten auf längeren Abschnitten.

Die Nutzung von Signalen von Detektoren zur lokalen Datenerfassung zur Fahrzeugwiedererkennung bietet den Vorteil, dass die Sensoren ohnehin schon vorhanden sind. Investitionen in die Sensor-Infrastruktur können entsprechend gering gehalten werden. Agg-

regierte Einzelfahrzeuggrößen, wie die Fahrzeuglänge, sind lokal und zunehmend auch in Verkehrszentralen zur weiteren Verarbeitung verfügbar. Detaillierte Fahrzeugsignaturinformationen bieten ein größeres Potential zur Wiedererkennung, können beim aktuellen Stand der Steuergeräte, der Kommunikationsinfrastruktur und der Zentralen jedoch nicht ohne größeren technischen und administrativen Aufwand genutzt werden. Ferner sind die Sensoren in städtischen Netzen meist nahe der LSA Haltelinie installiert, so dass die Muster häufig durch Brems- und Beschleunigungsvorgänge verzerrt werden. Bei der Signaturwiedererkennung können Fahrzeuge - anders als bei der Kennzeichenerkennung - meist nicht eindeutig wiedererkannt werden ("Ein VW Golf ist ein VW Golf"). In stark vermaschten städtischen Netzen ist eine zuverlässige Signaturwiedererkennung vermutlich nur durch eine Berücksichtigung zusätzlicher Informationen, z.B. aus einem Verkehrsmodell, möglich.

Fahrzeugwiedererkennung mittels Kennzeichenerfassung

Eine in der Praxis relativ häufig eingesetzte Technologie zur Fahrzeugwiedererkennung ist die automatische Erkennung und Zuordnung von Fahrzeugkennzeichen. Dazu werden geeignete Kameras an verschiedenen Stellen im Netz installiert, die mittels Verfahren der automatischen Bildauswertung und Buchstabenerkennungsalgorithmen (*engl.: Optical Character Recognition, OCR*) Zeichenfolgen aus Kennzeichen extrahieren und als ID des Fahrzeuges speichern. Die extrahierten und in der Regel verschlüsselten Zeichenfolgen sind eindeutig. Bei einer korrekten Bildauswertung ist eine falsche Zuordnung also nicht möglich.

Die Erfassungsrate einer Kamera bezogen auf einen Querschnitt hängt von der Qualität der Bildverarbeitungshard- und -software ab. Mit derzeit eingesetzten, sorgfältig eingestellten Systemen lassen sich bezogen auf einen Fahrstreifen Erfassungsraten von deutlich über 50% erreichen [SPANGLER 2008].

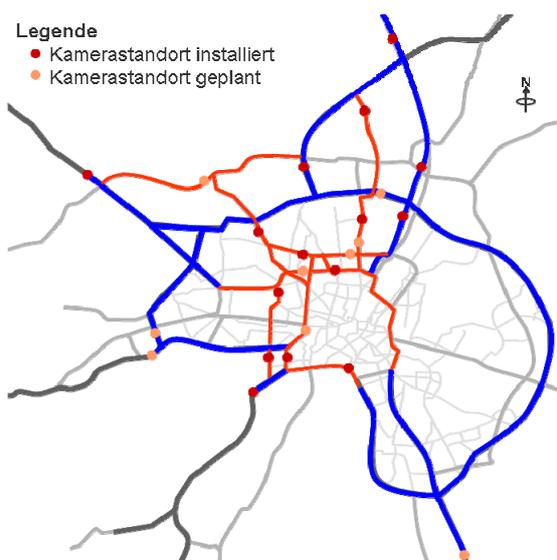


Abb. 2.4 Münchner Kennzeichenkamera-Messstellennetz [SPANGLER 2008]

Damit wäre bei mittleren und hohen Verkehrsstärken eine quasi kontinuierliche Beobachtung der Reisezeit möglich. Die Gesamtzahl der erfassten Fahrzeuge hängt jedoch vor allem davon ab, wie viele Fahrzeuge die betrachtete Route insgesamt durchfahren haben. Beispiele für erfolgreich operativ eingesetzte Systeme sind das "Section Control" System in Österreich zur Überwachung Routenbezogener Geschwindigkeiten (z.B. auf der A22 und der A1) und ein Messstellennetz zum Monitoring der Verkehrsqualität auf verschiedenen Routen des Münchner Hauptstraßennetzes (siehe Abb. 2.4).

2.2.3 Ermittlung von Reisezeiten aus fahrzeuggenerierten Daten

Fahrzeuggenerierte Daten (*engl.: Floating Car Data, FCD*) sind Daten, die aus im Verkehr mitschwimmenden Fahrzeugen generiert werden. Im Gegensatz zur Fahrzeugwiedererkennung, wird die Orts-Zeit Information im Fahrzeug erhoben und von dort (ggf. vorverarbeitet) an eine Zentrale gesendet.

Bei fahrzeuggenerierten Daten kann es sich prinzipiell auch um über die Position hinausgehende Informationen handeln, z.B. Daten der fahrzeuginternen Sensorik wie Scheibenwischer, Scheinwerfer, Temperatur, ESP etc. Man spricht dann von erweiterten fahrzeuggenerierten Daten (*engl.: extended FCD, xFCD*; siehe beispielsweise HUBER [2001]). Im Folgenden wird unter fahrzeuggenerierten Daten ausschließlich die Verwendung von Positionen und zugehörigen Zeitstempeln verstanden.

Die Nutzung von FCD zur Ermittlung von Reisezeiten ist seit den 1980er Jahren Gegenstand der Forschung. Beispielsweise wurden zwischen 1996 und 2001 im Projekt "MOTIV" etwa 100 Testfahrzeuge im Großraum München zur Ermittlung von Reisezeiten eingesetzt.

Obleich der Nutzen von FCD als Ergänzung zur herkömmlichen lokalen Datenerfassung offensichtlich ist, konnten nur wenige Systeme den Schritt von der Forschung hin zum produktiven System vollziehen. Als maßgeblich hierfür sind vor allem die Kosten für die Datenübertragung, die Anzahl der potentiell zur Verfügung stehenden Fahrzeuge insgesamt und die Angst um die Privatsphäre des Individuums zu sehen.

Aus praktischer Sicht eignen sich vor allem Flottenfahrzeuge als FCD Lieferanten. Diese haben typischerweise hohe Kilometerleistungen und sind zu Dispositionszwecken mit der entsprechenden Hardware ausgestattet, die ihre Position an eine Zentrale übermitteln. FCD können daher als "Abfallprodukt" ohne zusätzliche Kosten für Hardware und Kommunikation generiert werden. Bei der Interpretation der Daten muss allerdings das möglicherweise nicht repräsentative Fahrverhalten (z.B. maximale Geschwindigkeit von LKW) berücksichtigt werden.

Fahrzeuggenerierte Daten aus Fahrzeugen des ÖPNV

Die Fahrzeugortung im öffentlichen Personennahverkehr (ÖPNV) im Rahmen von rechnergestützten Betriebsleitsystemen geschieht heutzutage meist noch über die Ortung an Haltestellen über ein Baken-System (häufig per Infrarot) entlang der Route. Die Ortung erfolgt in der Regel an den Haltestellen unter Berücksichtigung der Türöffnung. Da im ÖPNV üblicherweise die Fahrtroute festliegt, kann die Position im Straßennetz über eine logische Ortung über den zurückgelegten Weg von der letzten Haltestelle ermittelt werden. Zur Kommunikation mit der Zentrale kommen häufig schmalbandige analoge Datenfunksysteme zum Einsatz [LÖWE 2004].

Obwohl vielversprechend, ist bisher keine Nutzung von Positionsmeldungen von Fahrzeugen des ÖPNV zur Anwendung in einem online Verkehrsinformationssystem für den Individualverkehr bekannt. Das mag zum Teil daher rühren, dass Fahrzeuge des ÖPNV aufgrund spezieller Fahreigenschaften als nicht relevant für den MIV eingestuft werden, zum Teil aber auch daher, dass die ÖV Betriebe ungern Daten ihrer Fahrzeuge in die Öffentlichkeit bringen. Fahrzeuggenerierte Daten aus Buspositionsmeldungen wurden zum Beispiel von TANTIYANUGULCHAI UND BERTINI [2003] zur Analyse der Reisezeit auf einer stark frequentierten Route in Portland/Oregon (USA) genutzt. Dabei wurden Methoden zur Bereinigung der Busreisezeiten von Haltestellenaufhalten entwickelt und eingesetzt.

Fahrzeuggenerierte Daten aus Taxipositionsmeldungen

Aufgrund der oben genannten Eigenschaften eignen sich Taxifloten gut als FCD Lieferanten. Es existieren bereits verschiedene Verfahren und Systeme in operativem oder semi-operativem Betrieb. Beispiele sind Berlin und Nürnberg (Systeme des Deutschen Luft- und Raumfahrtzentrums, siehe beispielsweise SCHÄFER [2005]), das FLEET System in Wien (System von Arsenal Research, LINAUER ET AL. [2006]), das OnVLS in Graz und NOSTRADAMUS in Guangzhou (Systeme der TU München, SCHIMANDL UND LEONHARDT [2008]).

Taxipositionsmeldungen stellen eine wichtige Datenquelle in den im Rahmen dieser Arbeit betrachteten Testfeldern München und Graz dar. Abb. 2.5 zeigt nach Status gefilterte, also potentiell nutzbare Taxipositionsmeldungen vom 11. März 2008 im Norden von München (links). Die unterschiedlichen Abdeckungsgrade der einzelnen Netzbereiche sind deutlich erkennbar. Rechts sind nutzbare Taxifahrten vom 13. März 2007 in Graz dargestellt. Hier wurden die zusammenhängenden Positionsketten durch lineare Verbindungen visualisiert.

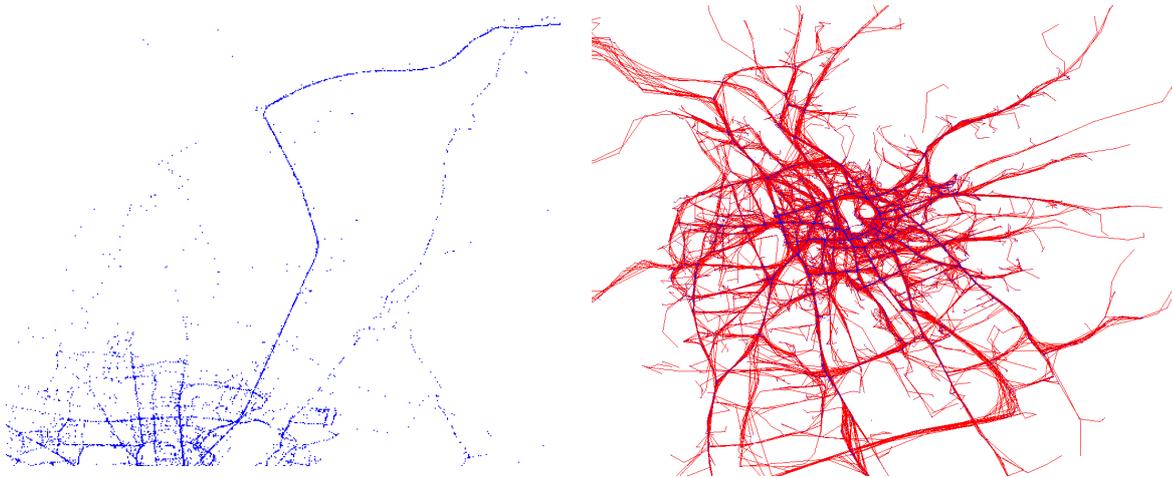


Abb. 2.5 Darstellung der potentiell nutzbaren Taxifahrten vom 13.03.2007. Links: München.
Rechts: Verbindung der nutzbaren Taxipositionen im Grazer Hauptstraßennetz.

Die Taxis melden ihre Position, abhängig vom Fahrtzweck, etwa alle 20 bis 180 Sekunden. In Abb. 2.6 dargestellt ist eine Fahrt in München auf der Ingolstädter Straße stadtauswärts. Die Position wird etwa alle zwei Minuten an die Dispositionszentrale übermittelt.



Abb. 2.6 Darstellung einer Taxifahrt

Das zum Einsatz in Graz und München entwickelte Verfahren zur Ermittlung von Kanten oder Routenreisezeiten ist in Abb. 2.7 dargestellt.

In beiden Netzen variieren die Abdeckungsgrade je nach Route und es liegen Werte unregelmäßig im Abstand von Minuten bis Stunden vor. Die empfangenen Taxipositionsmeldungen werden nach dem Fahrzeug-Status gefiltert.

Dabei werden die nicht zielgerichteten Fahrten (z.B. "suchend", "am Standplatz") im Weiteren nicht berücksichtigt. Nach einer Sortierung liegen zusammenhängende Positionsketten zur weiteren Verarbeitung vor.

Mittels Map Matching wird jede Fahrzeugposition auf das digitale Straßennetz referenziert. Dabei werden Kanten innerhalb eines zu definierenden Abstands (hier verwendet: 30 Meter) als potentiell befahrene Kante ("Kandidat") gekennzeichnet.

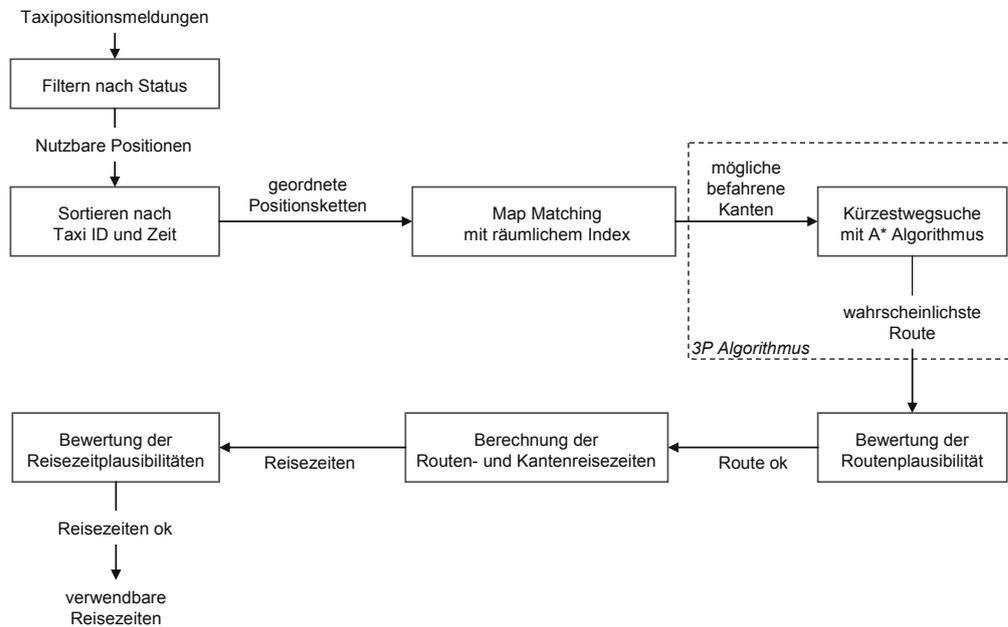


Abb. 2.7 Darstellung des Verfahrens zur Ermittlung von Reisezeiten aus Fahrzeugpositionsmeldungen

Da die Berechnung des Abstandes d für jede Kante j bei großen Netzen und einer großen Fahrzeugflotte sehr rechenintensiv wird, wird das Netz in einem Vorverarbeitungsschritt räumlich indiziert. Dabei wird das gesamte Untersuchungsgebiet zunächst in Felder gleicher Größe eingeteilt und für den Start- und den Endknoten jeder Kante werden die Indizes (i_{max} , i_{min} , j_{max} , j_{min}) ermittelt. Die Distanzen zur aktuellen Taxiposition werden dann nicht mehr für jede Kante, sondern lediglich für die den Feldern (i, j) mit $(i_{Taxi} + 1 \geq i \geq i_{Taxi} - 1$ UND $j_{Taxi} + 1 \geq j \geq j_{Taxi} - 1)$ zugeordneten Kanten berechnet. Um keine potentiell korrekte Kante auszulassen, muss die Größe der Felder mindestens die zu erwartende bzw. maximal tolerierte Abweichung zwischen Kante und Fahrzeugposition sein. Tab. 2.1 zeigt einen Vergleich der für das Map Matching benötigten Rechenzeit auf einem Standard-PC (2.4GHz Prozessor, 2GB RAM) für verschiedene Netze, Fahrzeugpositionsmengen und Feldgrößen. Durch die räumliche Indizierung kann also eine deutliche Geschwindigkeitssteigerung erreicht werden.

	Anzahl der Fahrzeugpositionen	Naives Map Matching	Spatial Index	Größe der Felder
Graz/Österreich (Hauptstraßennetz, 3700 Kanten)	1000	~ 15 Sekunden	~ 0.3 Sekunden	120x120 m ²
	5000	~ 70 Sekunden	~ 1 Sekunde	120x120 m ²
Guangzhou/China (Teilnetz, 15500 Kanten)	1000	~ 80 Sekunden	~ 2.5 Sekunden	120x120 m ²
	5000	~ 350 Sekunden	~ 11 Sekunden ~ 0.5 Sekunden	120x120 m ² 60x60 m ²

Tab. 2.1 Dauer des Map Matching Vorgangs für verschiedene Matching Verfahren

Mittels Kürzestweg-Suche wird der wahrscheinlichste Weg durch die Kandidaten-Kanten ermittelt. Dabei liegt die Annahme zugrunde, dass Fahrzeuge den kürzesten möglichen Weg zwischen mehreren Stationen im Netz nutzen. Zur Auswertung der anfallenden Positionsmeldungen wurde der 3P-Algorithmus (3 Positionen-Algorithmus) entwickelt. Dabei werden in jedem Berechnungsschritt drei aufeinander folgende Positionsmeldungen betrachtet. Im ersten Berechnungsschritt (*Initialschritt*) wird der kürzeste Weg berechnet, der alle Fahrzeugpositionen enthält. In den folgenden Berechnungsschritten wird jeweils die mittlere Kante des vorhergehenden Berechnungsschrittes fix als Startlink für den aktuellen Berechnungsschritt genutzt. Dadurch bleiben zum einen die einzelnen Berechnungsschritte konsistent und zum anderen reduziert sich die Anzahl der notwendigen Berechnungsschritte. Das Vorgehen ist in Abb. 2.8 dargestellt.

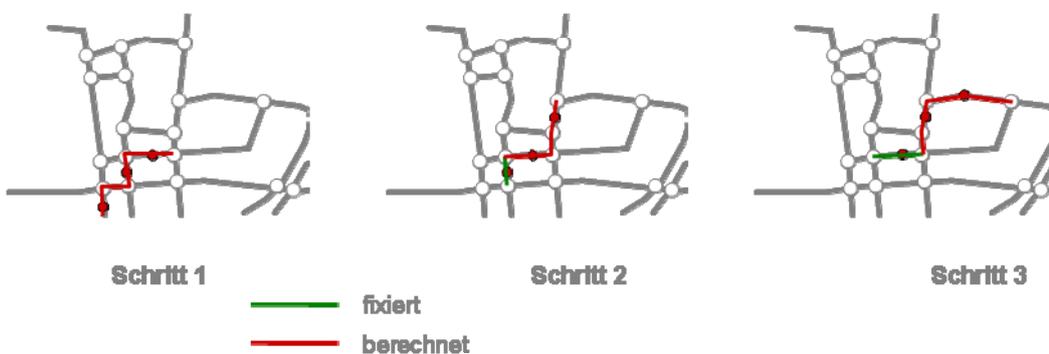


Abb. 2.8 Illustration der Berechnung der wahrscheinlichsten Route

Zur Routensuche kann prinzipiell jeder Kürzest-Weg Algorithmus verwendet werden. Aus Effizienzgründen wird mit dem A* ("A Stern") – Algorithmus ein informiertes Suchverfahren eingesetzt, was der uninformierten Suche (z.B. nach DIJKSTRA [1959]) bezüglich Rechenzeit überlegen ist. In Anhang 1 sind beide Algorithmen gegenübergestellt. Detaillierte Erläuterungen zur Problematik der Kürzestwegsuche sind beispielsweise in CHEN ET AL. [2005] zu finden. Zur effizienten Verarbeitung wird der Netzgraph in die so genannte Erweiterte Vorwärts Stern (EVS)-Struktur transformiert [ZILIASKOPOULOS UND MAHMASSANI 1996]. In der EVS Struktur besitzt jede Kante einen Zeiger zu ihren Folgekanten und erlaubt die implizite Berücksichtigung von Abbiegeverboten.

Die gefundene Route wird auf ihre Plausibilität und weitere Verwendbarkeit geprüft. Dazu wird die *Direktheit der Fahrt* (Verhältnis der Luftliniendistanz zur gefahrenen Route) ermittelt. Unterschreitet die Direktheit einen bestimmten Wert, so deutet dies auf eine fehlerhafte Routenrekonstruktion hin. Dies kann durch eine fehlerhafte Kartengrundlage, fehlerhaftes Map Matching durch ungenaue Positionsmeldungen oder ein "speziell" angepasstes Fahrverhalten (beispielsweise das Ignorieren von Linksabbiegeverboten oder die Nutzung der Fußgängerzone) verursacht werden.

Für die zurückgelegte Route kann nun die Reisezeit berechnet werden. Bei der Betrachtung von Teilrouten (z.B. einzelnen Kanten) werden die Reisezeiten interpoliert und geprüft ob sie in einem plausiblen Wertebereich liegen. Die als plausibel eingeschätzten Reisezeiten werden zur weiteren Verwendung gespeichert.

2.2.4 Nachbearbeitung der Einzelfahrzeugreisezeiten

Nachfolgend werden drei im Rahmen dieser Arbeit verwendete Methoden zur Filterung und Glättung der Einzelfahrzeugmessungen vorgestellt.

Überholfilter: Wenn ein Fahrzeug von einem anderen Fahrzeug überholt wird, dann wird das überholte Fahrzeug herausgefiltert. Annahme: Die Reisezeit des überholenden Fahrzeugs entspricht eher der realisierbaren Geschwindigkeit (weitere Erläuterungen zum Prinzip siehe ROBINSON UND POLAK [2006]).

Ähnlichkeitsvergleich aufeinanderfolgender Messwerte: Werden drei Messwerte innerhalb eines vorzugebenden Zeitintervalls Z erfasst und weicht der mittlere um mehr als ein zu definierender Schwellenwert S von beiden ab, so wird der mittlere Wert als unplausibel herausgefiltert [SPANGLER 2008]. Bei der Anwendung auf Realdaten haben sich $Z = 20$ Minuten und $S = 2$ Minuten als sinnvolle Werte herausgestellt.

Medianfilter: Aus innerhalb eines Aggregierungsintervalls erfassten Reisezeiten wird der Median als robuster Mittelwert berechnet (verwendetes Aggregierungsintervall: 15 Minuten).

Die Filter basieren alle auf einer gemeinsamen Betrachtung von zeitlich nah beieinander liegenden Reisezeitbeobachtungen. Dies ist bei der Verwendung von Taxipositionsmeldungen in aller Regel nicht gegeben, so dass am Ende für ein Zeitintervall meist und wenn überhaupt nur eine Einzelfahrzeugreisezeit zur Verfügung steht. Eine weitere Glättung der Daten, beispielsweise mittels exponentieller Glättung, wird aufgrund der Zufälligkeit der Beobachtung nicht durchgeführt.

2.3 Empirische Analyse der verfügbaren Daten

Im Rahmen dieser Arbeit werden die zwei Testgebiete Graz und München betrachtet. Beide Gebiete weisen jeweils unterschiedliche Datenverfügbarkeiten auf.

2.3.1 Testfelder und Datenquellen

Graz

In Graz liegen Daten lokaler Messstellen (aus Siemens CONCERT, insgesamt etwa 320 Messquerschnitte und Taxipositionsmeldungen der *Taxi878* Taxizentrale (i.d.R. etwa 160 Fahrzeuge

online) vor. Die Daten sind über das im Auftrag der Stadt Graz entwickelte Online Verkehrslagesystem (OnVLS) Graz verfügbar. Das OnVLS wurde und wird im Rahmen der Projekte „Integriertes Verkehrsmanagement (IVM)“ Graz (TRENDSETTER / EU FP5) und TRANSFER („Intermodales Reiseinformationssystem mit Echtzeitauskunft und kombinierten Parkfahrkarten“, gefördert vom BMVIT) entwickelt. Das Grazer Hauptstraßennetz umfasst 3770 gerichtete Kanten, die Gesamtnetzlänge des Hauptstraßennetzes beträgt etwa 450 Kilometer.

Für die Untersuchungen wird das Verfahren zur Prognose des Belegungsgrades an zwei Detektoren (EXERZIERPLATZ und ERZHERZOG) und zur Prognose der Reisezeit aus Taxipositionsmeldungen auf zwei Routen (KEPPLER-WEST und OPERNRING) eingesetzt (siehe Abb. 2.9 und Tab. 2.2).

Graz	Verwendungszweck
EXERZIERPLATZ Grazer Bundesstraße @ Exerzierplatzstraße (stadteinwärts)	Prognose des Belegungsgrades. Abhängige Variable: Belegungsgrad des lokalen Detektors "EXERZIERPLATZ". Unabhängige Variable: Belegungsgrade umliegender Detektoren.
ERZHERZOG Willhelm-Fischer-Allee @ Erzherzog-Johann-Allee (Richtung Westen)	Prognose des Belegungsgrades. Abhängige Variable: Belegungsgrad des lokalen Detektors "ERZHERZOG". Unabhängige Variable: Belegungsgrade umliegender Detektoren.
KEPPLER-WEST Kepplerstraße @ Grazer-Bundesstraße bis Kepplerstraße @ Bahnhofgürtel (Richtung Westen), Routenlänge: 1700 Meter	Prognose der Reisezeit. Abhängige Variable: Reisezeiten aus Taxipositionsmeldungen auf der Route "KEPPLER-WEST". Unabhängige Variable: Belegungsgrade von auf der Route und stromaufwärts gelegenen Detektoren.
OPERNRING Willhelm-Fischer-Allee @ Glacisstraße bis Joanneumring @ Wielandgasse (Richtung Westen), Routenlänge: 1200 Meter	Prognose der Reisezeit. Abhängige Variable: Reisezeiten aus Taxipositionsmeldungen auf der Route "OPERNRING". Unabhängige Variable: Belegungsgrade von auf der Route und stromaufwärts gelegenen Detektoren.

Tab. 2.2 Untersuchungsgebiet Graz, Untersuchungsgegenstände der einzelnen Lokationen

München

In München liegen Verkehrsstärkedaten, sowie Reisezeiten aus Fahrzeugwiedererkennung mit Kennzeichenkameras und aus Taxipositionsmeldungen vor. Kameras zur Kennzeichenerfassung sind und werden zur Überwachung der Verkehrsqualität auf Basis von Reisezeiten seit 2006 auf ausgewählten Routen installiert (Abschnitt 2.2.2). Das hier entwickelte Prognoseverfahren wird anhand von Daten einer der ersten ausgestatteten Routen (INGOLSTÄDTER, Abb. 2.10 und Tab. 2.3) erprobt. Zur Nutzung als unabhängige Variable liegen Verkehrsstärkedaten der lokal verkehrsabhängigen LSA auf der Ingolstädter Straße vor.

Auf zwei weiteren Routen (B2R und A9) werden ebenfalls Reisezeiten mit Fahrzeugwiedererkennung erfasst. Diese werden zur qualitativen Verifizierung der mittels Taxipositionsmeldungen ermittelten Reisezeiten verwendet (Abschnitt 2.3.2). In Tab. 2.2 und Tab. 2.3 sind die Untersuchungsgegenstände und die betrachteten Variablen der betrachteten Routen und Detektoren für Graz und München dargestellt.

München	Verwendungszweck
INGOLSTÄDTER Petuelring @ Sankt-Blasien-Straße bis Ingolstädter Straße @ Neuherbergerstraße (stadtauswärts), Routenlänge: 5300 Meter	Prognose der Reisezeit. Abhängige Variable: Reisezeiten aus Fahrzeugwiedererkennung auf der Route "INGOLSTÄDTER". Unabhängige Variable: Verkehrsstärken von auf der Route gelegenen Detektoren.
B2R Petuelring @ Sankt-Blasien-Straße bis A9 @ AS Freimann (stadtauswärts), Routenlänge: 5800 Meter	Visueller Vergleich von Reisezeiten aus Taxipositionsmeldungen und Fahrzeugwiedererkennung zur Verifikation der Methode der Reisezeitermittlung aus Taxireisezeiten und zur Überprüfung deren Repräsentativität.
A9 A9 @ AS Freimann bis A9 @ AS Garching-Nord (stadtauswärts), Routenlänge: 7600 Meter	Visueller Vergleich von Reisezeiten aus Taxipositionsmeldungen und Fahrzeugwiedererkennung zur Verifikation der Methode der Reisezeitermittlung aus Taxireisezeiten und zur Überprüfung deren Repräsentativität.

Tab. 2.3 Untersuchungsgebiet München, Untersuchungsgegenstände der einzelnen Lokationen

Übersicht



Abb. 2.9 Testgebiet Graz



Abb. 2.10 Testgebiet München

2.3.2 Aussagekraft der erhobenen Reisezeiten

Die aus Taxipositionsmeldungen und Fahrzeugwiedererkennung ermittelten Reisezeiten sollen die Trainingsbasis für ein Mustererkennungsverfahren zur Reisezeitprognose bilden. Daher ist die Validität und Repräsentativität der Daten äußerst wichtig.

Erster Schritt ist die Filterung der Reisezeiten mit den in Abschnitt 2.2.4 beschriebenen Filterverfahren. Abb. 2.11 zeigt die ungefilterten (grau) und die gefilterten (rot) Reisezeiten für die Routen "INGOLSTÄDTER". Die resultierenden Reisezeiten erscheinen plausibel und repräsentativ. Die effektive Anwendung der Filter ist nur bei ausreichenden Beobachtungshäufigkeiten möglich. Der Überholfilter und der Ähnlichkeitsvergleich können nur reagieren, wenn die Fahrzeuge in ausreichend nahem zeitlichen Abstand aufeinander folgen. Die Medianbildung macht nur Sinn, wenn mehr als eine Reisezeit im betreffenden Intervall vorliegt.

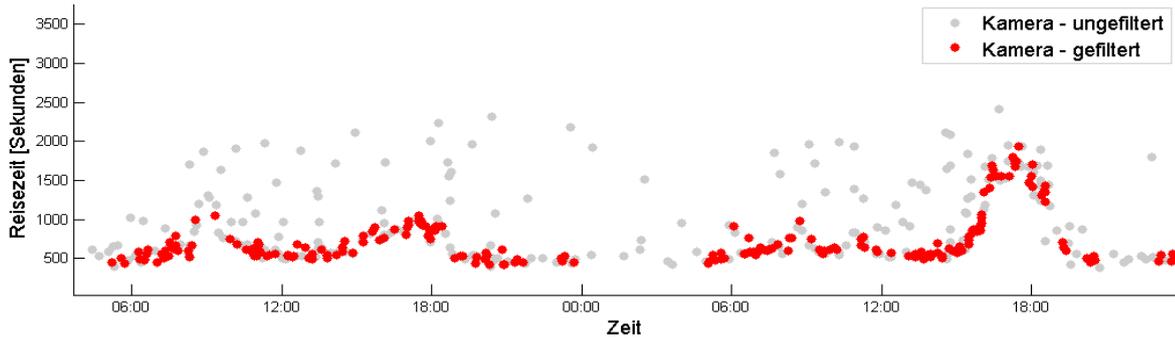


Abb. 2.11 Darstellung der Filterwirkung bei Anwendung auf mittels Kennzeichenerfassung ermittelten Reisezeiten (5.-6.09.2007/Petuelring – Ingolstädter Straße/München)

Vergleich von Taxireisezeiten und Reisezeiten aus Kennzeichenwiedererkennung

Häufig stehen ausschließlich relativ seltene Taxipositionsmeldungen zur direkten Ermittlung der Reisezeit zur Verfügung, so dass die Filter nicht verwendet werden können. Die so ermittelten ungefilterten Einzelfahrzeugreisezeiten stellen trotz einer jeweils korrekten Ermittlung nicht zwingend eine repräsentative und plausible Reisezeit dar. Zur Überprüfung der Repräsentativität wurden die Reisezeiten aus Taxipositionsmeldungen mit mittels Fahrzeugwiedererkennung erhobenen und gefilterten Reisezeiten auf zwei Routen im Münchner Straßennetz verglichen ("A9" in Abb. 2.12 und "B2R" in Abb. 2.13). Die Ergebnisse beider Methoden stimmen augenscheinlich gut überein, die Reisezeit aus Taxireisezeit kann hier bei Vorliegen als repräsentativ betrachtet werden.

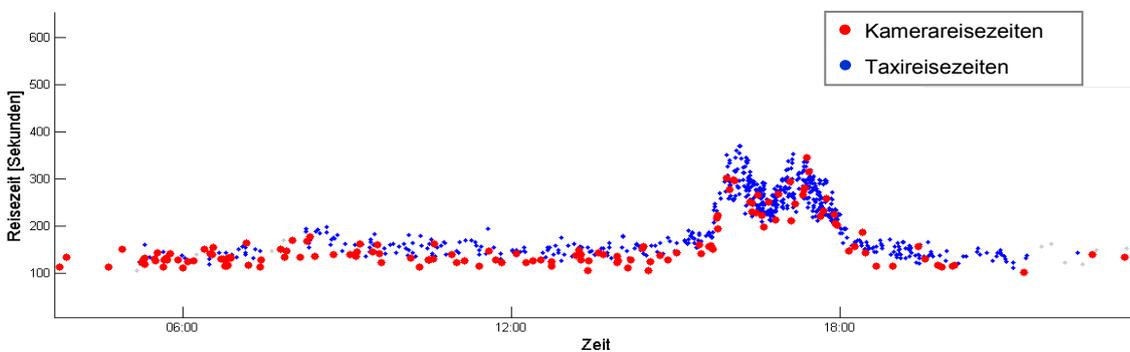


Abb. 2.12 Reisezeiten auf der A9

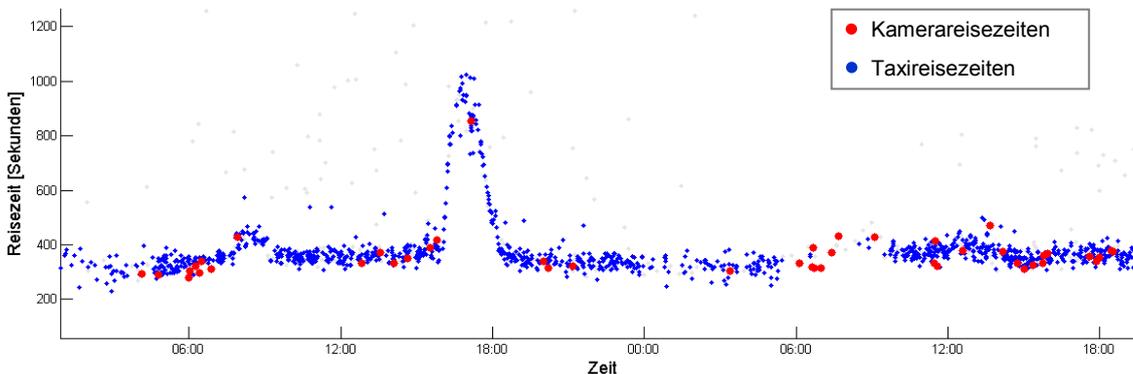


Abb. 2.13 Reisezeiten zwischen Petuelring und Freimann (B2R).

Auswirkung unterschiedlicher Routenlängen

Es ist anzunehmen, dass die Repräsentativität einer Einzelfahrzeugreisezeit mit abnehmender Routenlänge abnimmt, da Haltevorgänge an LSA bei kurzen Reisezeiten relativ betrachtet stärker ins Gewicht fallen. So gesehen wäre das Ziel, möglichst lange Routen zu wählen. Jedoch nimmt die Anzahl der Beobachtungen mit der Länge der Route ab, da nur eine Messung vorliegt, wenn die Route von einem Taxi komplett durchfahren wird. Somit wären kürzere Routen zu bevorzugen.

Zur Illustration ist der Zusammenhang zwischen Routenlänge, durchschnittlicher Anzahl von Beobachtungen pro Tag und der möglichen relativen Abweichung von zwei kurz aufeinander folgend erhobenen Einzelfahrzeugreisezeiten beispielhaft für eine Route in Graz dargestellt (siehe Abb. 2.14). Dabei wurde vereinfachend angenommen, dass der Unterschied zwischen zwei kurz aufeinander folgenden Fahrzeugen maximal einen Halt auf der gesamten Route beträgt (das vordere Fahrzeug kann eben noch die Grünphase nutzen). Bei einer Umlaufzeit von 90 Sekunden und einer effektiven Grünzeit von 30 Sekunden würde der maximale absolute Unterschied der Reisezeiten 60 Sekunden betragen. Die typische relative Abweichung wird berechnet aus dem maximalen Unterscheid (60 Sekunden) bezogen auf den Median der Reisezeiten auf der Route über den gesamten Tag. Bei hoher Verkehrsnachfrage und resultierenden hohen Reisezeiten reduziert sich die relative Abweichung entsprechend.

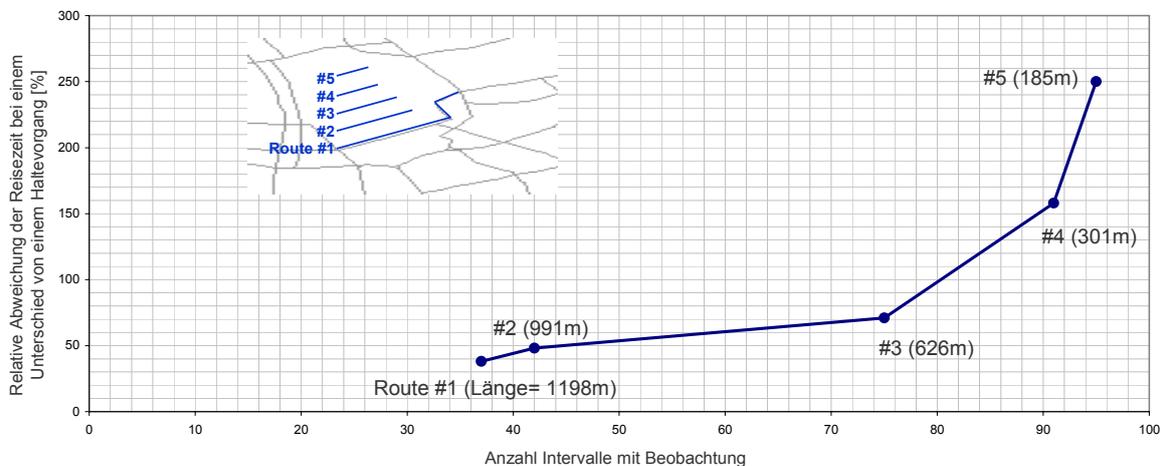


Abb. 2.14 Zusammenhang zwischen Routenlänge, durchschnittlicher Anzahl von Beobachtungen pro Tag und möglicher relativen Abweichung von zwei aufeinander folgend erhobenen Einzelfahrzeugreisezeiten

Insgesamt unterliegt die Festlegung der Routenlänge widersprüchlichen Kriterien: Eine Reduktion der möglichen relativen Abweichung ist stets mit einer Reduktion der Beobachtungsanzahl verbunden und umgekehrt. Somit wird die Festlegung der Routenlänge im Prinzip zu einer Pareto-Optimierung mit den Zielen "möglichst repräsentative Reisezeiten" und "möglichst viele Beobachtungen".

Im praktischen Einsatz sollte im Einzelfall geprüft werden, welche Routenfestlegung sinnvoll ist. Diese hängt auch vom Anwendungszweck ab. Für eine möglichst schnelle Störungserkennung wären etwa kurze Routen günstiger, während möglichst genaue Reisezeiten größere Routenlängen erfordern. Gleichzeitig muss die Routenlänge sinnvoll begrenzt werden, um überhaupt verkehrstechnisch relevant zu sein (die Einteilung der Routen sollte sich an wichtigen Entscheidungspunkten im Netz orientieren). In der vorliegenden Arbeit werden die Reisezeiten als historische Stichprobenmessung verwendet. Da hier von relativ großen Archivierungszeiträumen (mehrere Monate) ausgegangen wird und das Verfahren die Erfassung zu jedem Zeitpunkt nicht erfordert, erscheint es sinnvoll, die Routen im Zweifelsfall eher länger zu wählen.

2.3.3 Korrelationen und Dynamik verkehrlicher Kenngrößen

Eignung lokaler Verkehrskenngrößen als Prädiktorvariable

Lokale Verkehrskenngrößen eignen sich aufgrund ihrer quasi ständigen zeitlichen Verfügbarkeit prinzipiell gut als Prädiktorvariablen, da eine auf ihrer Basis zu generierende Prognose zu jedem Zeitpunkt ermittelt werden kann. Neben der Verfügbarkeit ist jedoch auch die Aussagekraft der Kenngröße zur Prognose der Kenngröße selbst oder der Reisezeit in der näheren Umgebung relevant. Dabei sind insbesondere Zeitbereiche mit hohen Verkehrsbelastungen und Überlastungen relevant. Innerorts stehen meist Einzel-Induktivschleifen Detektoren zur Verfügung, häufig im direkten Einflussbereich von Lichtsignalanlagen. Typischerweise werden dabei Verkehrsstärken q und Belegungsgrade b erfasst.

Bei Vorliegen beider Kenngrößen sollte b als beschreibende Variable gewählt werden, da sie auch Informationen über Stauzustände liefern kann. Abb. 2.15 zeigt b - q -Diagramme für zwei Detektoren im Grazer Straßennetz. Der Zusammenhang unterscheidet sich von Messstelle zu Messstelle und hängt natürlich auch von den beobachteten Belastungen ab. Während im Bereich niedriger Belegungsgrade ein guter linearer Zusammenhang festgestellt werden kann, sind in Bereichen hoher Belegungsgrade aufgrund beschränkter Kapazitäten keine signifikanten Änderungen (links) beziehungsweise sogar fallende Werte in der Verkehrsstärke zu erkennen (rechts, überstauter Zustand in dem die "normale" Kapazität nicht mehr realisiert werden kann). In jedem Fall lässt sich die Verkehrsstärke relativ eindeutig aus dem Belegungsgrad bestimmen. Umgekehrt gilt dies nicht, da im überlasteten Bereich – etwa durch Rückstau aufgrund einer Überlastung der stromabwärtigen LSA – die Verkehrsstärke einen relativ konstanten Wert annimmt (Kapazität) und bei schwerwiegenden Störungen sogar wieder abnimmt. Umgekehrt steigt die Belegung auch im Überlastbereich noch an oder bleibt im ungünstigsten Fall konstant – sie nimmt jedoch bei einer die Kapazität übersteigenden Nachfrage nicht ab.

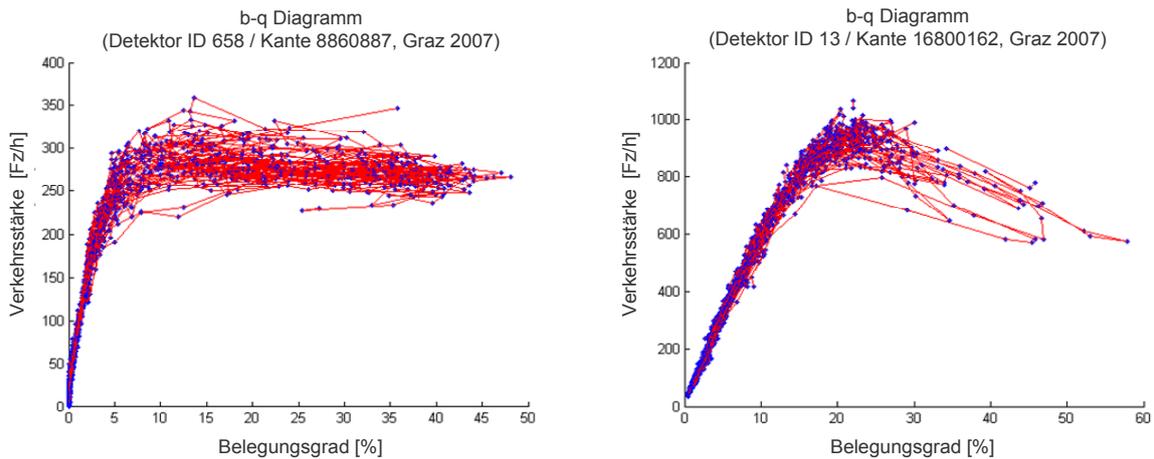


Abb. 2.15 Belegungsgrad-Verkehrsstärke Diagramme (basierend auf 15-Minutenwerten)

Je nach Erfassungstechnologie, Zentralenausstattung und Primärnutzung liegen teilweise ausschließlich Verkehrsstärkemessungen vor, etwa für die hier betrachtete Route "INGOLSTÄDTER". Bei den Detektoren handelt es sich ausschließlich um "halteliniennahe" Detektoren in den Zufahrten der LSA. Die Verkehrsstärke eignet sich aus den oben genannten Gründen nicht optimal als Prädiktor für die Reisezeit. Abb. 2.16 zeigt die Problematik. Der Zusammenhang zwischen der Reisezeit und der Verkehrsstärke an einem Detektor auf der entsprechenden Route ist alles andere als eindeutig (links) und auch bei Berücksichtigung der Zeit als erklärende Variable ist eine eindeutige Beschreibung in aller Regel nicht möglich (rechts). Dies verdeutlicht, dass es schwierig oder vielleicht nicht möglich sein dürfte, alleine aus der Verkehrsstärke die Reisezeit zu schätzen oder gar zu prognostizieren.

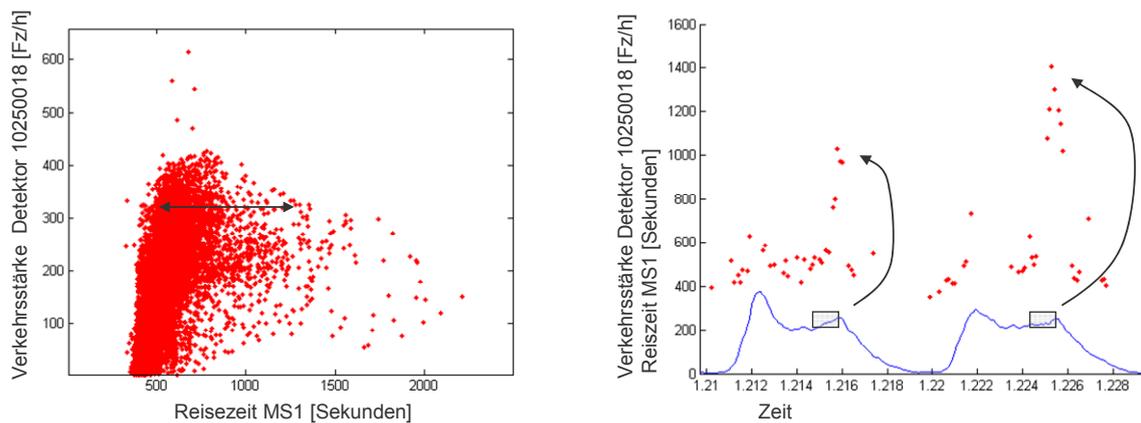


Abb. 2.16 Schätzung und Prognose von Reisezeiten aus lokalen Verkehrsstärkemessungen (München, Reisezeit (rot) und Verkehrsstärke (blau) eines LSA-Detektors auf der Route "INGOLSTÄDTER")

Kalenderabhängigkeit verkehrlicher Kenngrößen

Verkehrliche Kenngrößen unterliegen aufgrund wiederkehrender Nachfragestrukturen Tageszeit- und Wochentag-abhängigen Mustern. Je eindeutiger diese Muster sind, desto ein-

facher ist eine Schätzung und Prognose auf Basis von Kalendermerkmalen. Abb. 2.17 zeigt zwei Beispiele dafür, dass Tageszeit und Tagesklasse für eine Schätzung und Prognose nicht ausreichen. Links dargestellt ist ein Ausschnitt der Reisezeit-Zeitreihe für INGOLSTÄDTER. In blau sind die tatsächlichen Messungen dargestellt, in grau hinterlegt ist die typische Reisezeitganglinie in Abhängigkeit von Wochentag und Tageszeit. Teilweise stimmen beide gut überein, aber die außergewöhnlich hohe Reisezeit von über 1400 Sekunden kann, da sie eben außergewöhnlich ist, nicht durch die typische Ganglinie abgebildet werden. Rechts dargestellt sind tageszeitabhängige Belegungsgrade, wobei die Messwerte für verschiedene Tagesklassen farblich kodiert sind. Die typischen wochentagabhängigen Muster sind deutlich zu erkennen, jedoch herrschen insbesondere zu den verkehrlich interessanten Zeiten (Abendspitze mittwochs, vorgezogene Abendspitze freitags) deutliche Streuungen innerhalb der Klassen.

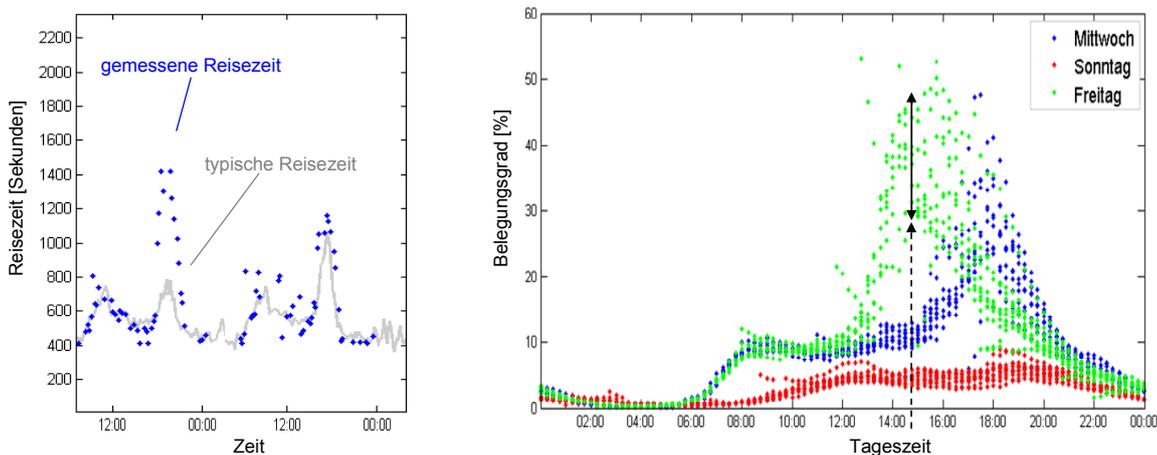


Abb. 2.17 Zeitabhängigkeit der Reisezeit (links, INGOLSTÄDTER) und des Belegungsgrades (rechts, ERZHERZOG)

Dynamik verkehrlicher Kenngrößen

Kontinuierlich vorliegende Verkehrsdaten werden meist als Zeitreihe, also in Form zeitlich aufeinanderfolgender Messwerte repräsentiert. Für Prognosefragen, insbesondere für sehr kurze Prognosehorizonte, ist die Intervall-zu-Intervall Dynamik der Zeitreihe eine wichtige Größe. Ist diese gering, d.h. unterscheiden sich aufeinander folgende Werte nur wenig, so liegen Trivialprognosen ($y(t+dt) = y(t)$) in der Regel richtig und es können sehr einfache Modelle zur Prognose eingesetzt werden.

Abb. 2.18 zeigt Belegungsgrade (y -Achse) als Funktion des jeweiligen Belegungsgrades 15 Minuten früher (x -Achse) am selben Detektor. Es ist deutlich zu erkennen, dass eine eindeutige Zuordnung des Belegungsgrades "15 Minuten später" (Prognose) schwerlich möglich ist. Die Betrachtung von mehreren aufeinander folgenden Messwerten als Muster kann die Prognose möglicherweise verbessern (links). Es ist jedoch zu erkennen, dass die Prognose

einer Zeitreihe aus sich selbst heraus problematisch ist. Das gilt insbesondere bei hohen Belegungsgraden, also relevanten Situationen, in denen selbst in den aggregierten und geglätteten Daten häufig sprunghafte Änderungen zu beobachten sind.

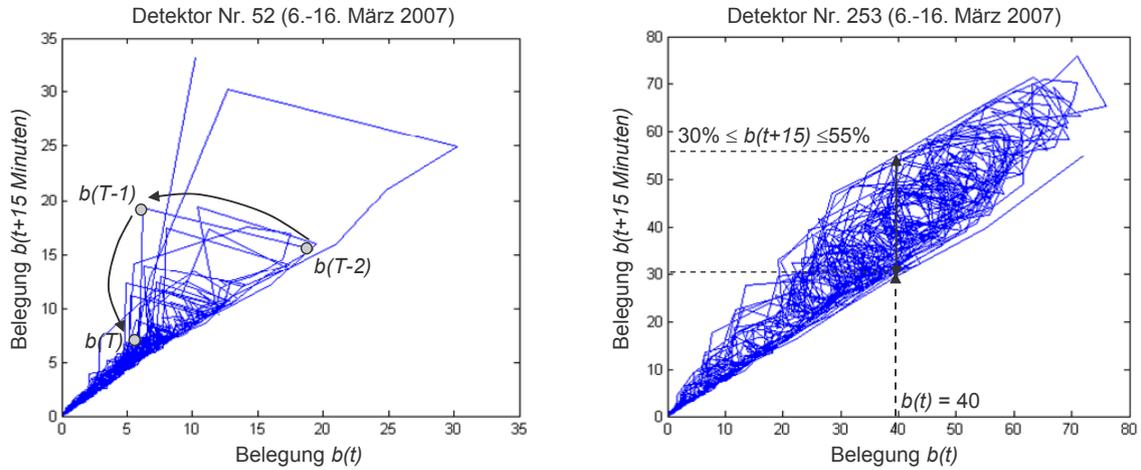


Abb. 2.18 Darstellung des Belegungsgrades als Funktion des Belegungsgrades 15 Minuten früher am gleichen Detektor

Zusammenhang von Verkehrskenngrößen an unterschiedlichen Erfassungsorten

Die Verkehrskenngrößen an unterschiedlichen Erfassungsorten, beispielsweise an zwei Detektoren auf zwei Kanten im Straßennetz, weisen mehr oder weniger starke Zusammenhänge auf. Werden diese über längere Zeit beobachtet und ist der Zusammenhang ausreichend eindeutig, so kann die eine Größe aus der anderen geschätzt werden. Mögliche Anwendungsszenarien können die Schätzung der Werte eines ausgefallenen Detektors oder die Bildung einer hypothetischen Zeitreihe zur Plausibilisierung der Detektordaten sein.

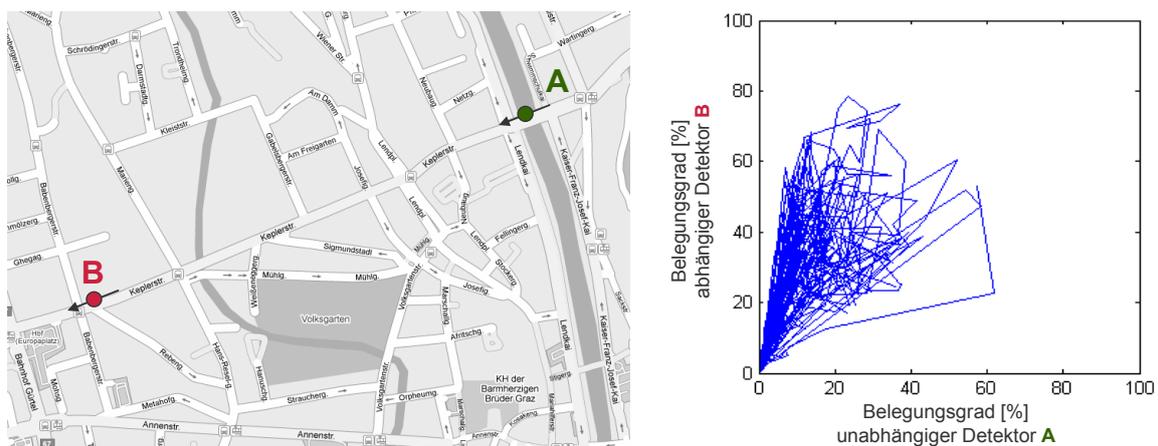


Abb. 2.19 Darstellung des Belegungsgrades eines Detektors B als Funktion des Belegungsgrades eines stromaufwärts gelegenen Detektors A.

Abb. 2.19 zeigt die Abhängigkeit des Belegungsgrades eines Detektors B als Funktion des Belegungsgrades eines stromaufwärts gelegenen Detektors A auf der Keplerstraße in Graz über einen Zeitraum von zehn Tagen. Beide sind, wenig überraschend, positiv korreliert. Jedoch ist der Zusammenhang insbesondere in Bereichen hoher Belegungsgrade nicht sehr eindeutig, so dass eine Schätzung der Belegung $b_B(t) = f(b_A(t))$ nicht sehr zuverlässig sein dürfte (der Korrelationskoeffizient für dieses Beispiel liegt bei 0,76). Eine stärkere Korrelation (und somit bessere Schätzung) könnte möglicherweise über die Berücksichtigung der Dynamik in Form der geschätzten Propagierungsgeschwindigkeit erreicht werden.

2.4 Fazit und weiteres Vorgehen

Als Grundlage zur Prognose verkehrlicher Kenngrößen stehen in städtischen Netzen häufig kontinuierlich erfasste lokale Verkehrsdaten und zufällig oder zeitlich über einen begrenzten Zeitraum beobachtete Daten zur Ermittlung von Reisezeiten zur Verfügung.

In Graz liegen Belegungsgrade lokaler Detektoren und Reisezeiten aus Taxipositionsmeldungen vor. In München stehen auf ausgewählten Routen Reisezeiten aus Kennzeichenerkennungssystemen und lokale Verkehrsstärken zur Verfügung.

Die auf Basis realer Daten aus den beiden Testfeldern durchgeführten exemplarischen visuellen Untersuchungen haben gezeigt, dass Zusammenhänge zwischen (1) lokalen Verkehrskenngrößen verschiedener Detektoren, zwischen (2) lokalen Verkehrskenngrößen und Reisezeiten und (3) innerhalb der Zeitreihen vorliegen. Es ist jedoch auch zu erahnen, dass einfache Ansätze, wie die einfache Regression, die Prognose einer Zeitreihe aus sich selbst heraus (autoprojektiv) oder die Prognose ausschließlich in Abhängigkeit von Kalendervariablen nicht ausreichen, um die komplexen Zusammenhänge in den Verkehrskenngrößen abzubilden.

Es wird davon ausgegangen, dass eine Betrachtung über einzelne lokale und zeitlich begrenzte Merkmale hinaus in Form räumlich-zeitlicher Muster unter Einbezug von Kalendermerkmalen zu einer verbesserten Schätzung und Prognose führen kann. Zusammengefasst werden die Merkmale zu einem Merkmalsvektor bzw. Muster, wobei die ausgewählten Merkmale die aktuelle Situation bezüglich der nahzukünftigen Entwicklung möglichst gut charakterisieren sollen.

Der Prozess der Identifikation und Nutzung dieser Muster zur Schätzung und Prognose kann unter dem Begriff (Daten-)Mustererkennung subsumiert werden. Im folgenden Kapitel werden die Mustererkennung im Allgemeinen und spezifische Methoden und Anwendungen der Mustererkennung zur Verkehrsprognose dargestellt.

3 Mustererkennungsmethoden und Anwendungen zur Verkehrsprognose

Auf Basis der definierten Anforderungen und der Eigenschaften der in der Praxis zur Verfügung stehenden Datenquellen werden in *Kapitel 3* die prinzipiell in Frage kommenden Klassen von Prognoseverfahren analysiert. Unter diesen, allgemein aus dem Bereich der "Datenmustererkennung" stammenden Verfahren, wird mit dem Instanzbasierten Lernen ein geeignetes ausgewählt. Wesentlich für die Auswahl sind dabei die Transparenz und die zu erwartende Robustheit des Verfahrens, sowie vor allem die Möglichkeit, relevante statistische Kenngrößen zur Prognosezuverlässigkeit direkt aus den Daten abzuleiten.

3.1 Grundlagen und Einordnung

Das Forschungsgebiet der Mustererkennung kann nach HAGEDORN ET AL. [1997] wie folgt definiert werden: "Gegenstand der Datenmustererkennung sind große, strukturierte Bestände numerischer, ordinal- oder nominalskaliertes Daten, in denen interessante, aber schwer aufzuspürende Zusammenhänge vermutet werden." ZITZELSBERGER [2006] liefert eine für diese Arbeit relevante Spezifizierung: "Die Mustererkennung [...] beinhaltet Verfahren zum automatisierten Erkennen und Auswerten von Mustern in ein- und mehrdimensionalen Signalen."

Die Mustererkennung oder Datenmustererkennung (engl.: *Data Mining*) beschäftigt sich also mit der Strukturierung und Auswertung großer Datenmengen. Der Antrieb ist die Vermutung von wertvollem, aber verstecktem Wissen in Datenarchiven, das durch Mustererkennungsverfahren aufgedeckt und nutzbar gemacht werden soll. Die Regression ist ein wichtiges Teilgebiet der Mustererkennung und beschäftigt sich mit der Schätzung und Prognose von numerischen Daten. Dazu werden auf Basis von Beobachtungen (Beispieldatensätzen) Zusammenhänge zwischen den zu prognostizierenden Größen (abhängige Variable) y und gemessenen unabhängigen Variablen X hergestellt.

$$y = f(X), \quad X = x_1, \dots, x_n$$

Beispiel

An einem Montag um 09:00 Uhr soll die Wartezeit an einer Lichtsignalanlage um 09:30 Uhr prognostiziert werden. Zur Verfügung stehen Messungen von stromaufwärts gelegenen lokalen Detektoren. Informationen zur Festzeitsteuerung liegen nicht vor und es gibt eine nicht detektierte Einmündung zwischen den Detektoren und der Lichtsignalanlage. In der Vergangenheit wurde die Wartezeit in unregelmäßigen Abständen mit Floating Cars beobachtet.

Die zukünftige Wartezeit (z.B. in 30 Minuten) kann unter der Annahme, dass sie zum Teil von der aktuellen Verkehrssituation in der Nähe der Lichtsignalanlage und zum Teil von typischen, wöchentlich wiederkehrenden Verkehrsmustern abhängt, prognostiziert werden. Mögliche Schritte wären die Suche nach ähnlichen Verkehrsmustern der Detektoren an den vergangenen Montagen für die Fälle, in denen 30 Minuten später die Wartezeit mit einem Floating Car beobachtet wurde. Aus den Mustern mit der größten Übereinstimmung könnte dann die Wartezeit in 30 Minuten geschätzt werden.

3.1.1 Stufen der Mustererkennung

Der Mustererkennungsprozess allgemein ist in Abb. 3.1 dargestellt. Die erfassten Rohdaten (hier: Verkehrsdaten) werden zunächst vorverarbeitet. Zur Vorverarbeitung gehören Plausibilisierung, Aggregation und Glättung, aber auch die Vorklassifizierung und die Segmentierung der Daten in sinnvolle Gruppen, sowie die Merkmalsreduktion. Bezogen auf das Ziel (hier: Prognose einer abhängigen Variable in Form einer Verkehrskenngröße, z.B. Reisezeit) werden relevante Merkmale in den Daten identifiziert und extrahiert. Der eigentliche Prozess der Mustererkennung ist die Bildung eines Modells aus einer Trainingsdatenbasis, also das Lernen von Eingabe-/Ausgabepaaren. Das gelernte Modell kann anschließend mit Testdaten oder mit im operativen Betrieb erfassten Daten gefüttert werden.

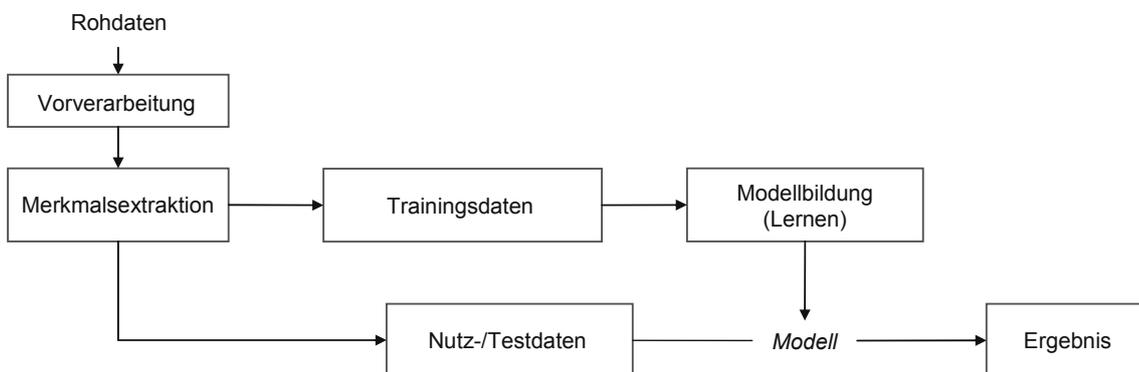


Abb. 3.1 Ablauf der Mustererkennung (in Anlehnung an NIEMANN [2003])

3.1.2 Verfahren der Mustererkennung

Es existieren verschiedene Verfahren der Mustererkennung zur Lösung von Regressionsproblemen. Häufig eingesetzte und hier näher betrachtete Ansätze sind:

Bei *regressionsanalytischen Verfahren* werden Zusammenhänge zwischen y und X explizit als Formel formuliert. Die Parameter werden durch Fehlerminimierung auf Grundlage einer Trainingsdatenbasis angepasst.

Künstliche Neuronale Netze sind flexible, aus einfachen Verarbeitungseinheiten (Neuronen) bestehende Funktionsapproximatoren. Aufgrund ihrer Struktur existieren effiziente Lernalgo-

rithmen zur optimalen Verknüpfung der Neuronen mit dem Ziel, den Zusammenhang zwischen X und y herzustellen.

Selbstlernende Expertensysteme sind hybride Verfahren, die Expertenwissen über Zusammenhänge zwischen X und y in Form prototypischer Regeln, Fallbeispielen oder Entscheidungsbäumen abbilden und diese auf Basis von Trainingsdaten optimieren.

Bei *Instanzbasierten Lernverfahren* (auch: *Parameterfreie Regression*) wird die Ähnlichkeit des aktuellen X zu den historischen Instanzen entsprechend eines vorgegebenen Ähnlichkeitsmaßes zur Laufzeit ermittelt und y aus den zugehörigen historischen abhängigen Variablen direkt auf Basis der Trainingsdaten approximiert.

3.1.3 Abgrenzung zu Verkehrsmodellen

Verkehrsmodelle bestehen aus erklärenden Modellen zur Schätzung der Verteilung der Fahrzeuge im Netz (Umlenkmmodelle), sowie der Abwicklung der Verkehrsnachfrage (Verkehrssimulationsmodelle). Randbedingungen sind dabei üblicherweise gemessene Verkehrsstärken an verschiedenen Stellen im Netz. Nachfragen werden in Form von Nachfragematrizen oder an den Netzrändern gemessenen Verkehrsstärken als Approximation der Nachfrage eingespeist. Prognosen können mit Verkehrsmodellen gerechnet werden, indem die Nachfrage prognostiziert und eingespeist wird.

Verkehrsmodelle erlauben die explizite Berücksichtigung von Änderungen im Angebot (z.B. Sperrung eines Fahrstreifens in Folge einer Baustelle) und von Verkehrsmanagementmaßnahmen (z.B. Änderung des Signalplans oder eine Umleitungsempfehlung). Daher sind sie zur simulativen Prognose der Wirkung von Szenarien geeignet. Neben der gerade in komplexen städtischen Netzen aufwändigen Kalibrierung ist die Prognose der Nachfrage an den Netzrändern ein zu lösendes Problem. Für bestimmte Prognosefragestellungen stellt sich daher die Frage, warum ein Verkehrsmodell überhaupt nachgeschaltet werden soll, wenn die interessierenden Größen mit Methoden der Mustererkennung auch direkt prognostizierbar sind. So stellt VORTISCH [2006, S. 109] fest:

"Vor diesem Hintergrund der Volldetektion (Anm.: direkte Erfassung von Reisezeiten, z.B. auf Basis von Mobilfunk) stellt sich die Frage nach dem weiteren Bedarf an modellgestützten Verfahren. Auf den ersten Blick erscheint spätestens bei der Aufgabe der Kurzfristprognose der Einsatz eines Verkehrsmodells unabdingbar. Wenn allerdings über einen längeren Zeitraum eine kontinuierlich vollständig erfasste Verkehrssituation zur Verfügung steht, können Verfahren z.B. auf der Basis künstlicher neuronaler Netze auch lernen, aus der aktuellen Situation die wahrscheinlich kommende Entwicklung vorherzusagen, ohne dazu ein explizites Verkehrsmodell zu benötigen. Wirklich benötigt werden Verkehrsmodelle aber dann, wenn von einer aktuellen Situation ausgehend die Wirkung verschiedener Maßnahmen vor-

hergesagt werden soll, wenn also eine maßnahmensensitive Kurzfristprognose berechnet werden soll."

Datenmusterbasierte Modelle und Verkehrsmodelle stehen nicht in Konkurrenz zueinander. Beide Ansätze stellen legitime Methoden zur Lösung von Prognoseproblemen dar und sollten entsprechend der Fragestellung ausgewählt werden. In vielen Anwendungen können sich beide Modellklassen sogar ergänzen (Ermittlung von Eingangsgrößen für die Mustererkennung durch Verkehrsmodelle [MAIER ET AL. 2008] oder Prognose von Nachfrageganglinien zur Verwendung in einem Verkehrsmodell [VORTISCH 2006]).

3.2 Klassifikation und Segmentierung der Datenbasis

Vor dem eigentlichen Mustererkennungsprozess können die vorhandenen Daten in einem Vorverarbeitungsschritt klassifiziert und/oder segmentiert werden. Klassifikation ist die überwachte Einordnung von Objekten in ein festgelegtes Klassenschema und dient der Einordnung der Daten in homogene Gruppen bezüglich a priori bekannter Merkmale. Bei der Segmentierung (*engl.: Clustering*) werden Objekte (meist Ganglinien) automatisch bezüglich vorgegebener Kriterien unüberwacht eingeteilt. Die Segmentierung dient vor allem der Reduktion der Datenbasis und der Hervorhebung von typischen Mustern.

Häufig werden beide Schritte durchgeführt. So stellt WILD [1996] seinem Verfahren zur Erkennung und Prognose von Ganglinien eine Klassifizierung in Tagesgruppen und eine automatische Segmentierung innerhalb der Tagesgruppen voraus (siehe Abb. 3.2). Ähnlich wird bei dem in der Verkehrsmanagementzentrale (VMZ) Berlin eingesetzten Verfahren VISUM-Online vorgegangen [VORTISCH 2006].

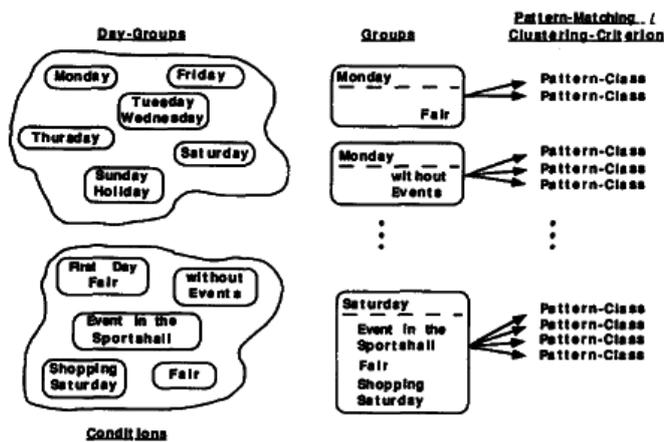


Abb. 3.2 Klassifizierung und Segmentierung der Ganglinienbasis [WILD 1996]

3.2.1 Klassifikation nach Kalendermerkmalen

Verkehrskenngrößen sind abhängig von der Verkehrsnachfrage und diese folgt mehr oder weniger stark ausgeprägten kalenderabhängigen Mustern (Montag bis Freitag um 08:00 zur Arbeit, jeden zweiten Samstag um 15:00 zum Fußballspiel, etc.). Die Klassifikation der Verkehrsdaten in möglichst homogene Gruppen bezüglich a priori bekannter, zeitlich eindeutiger Kalendermerkmale erscheint daher sinnvoll. Typische Kalendermerkmale sind Wochentage, Feiertage, Lage des Tags (z.B. vor/nach Feiertag, Brückentage, Schulferien) und verkehrlich wirksame Veranstaltungen (z.B. Fußballspiel).

Auf Basis von nach Kalendermerkmalen klassifizierten Zeitreihen kann eine historienbasierte Prognose ohne Berücksichtigung von aktuellen Verkehrsdaten berechnet werden (z.B. Geschwindigkeit an einem typischen Dienstag um 09:00 Uhr). Dazu werden aus den einer Klasse zugeordneten Zeitreihen Referenzganglinien gebildet (Klassenmittelpunkt, z.B. arithmetischer Mittelwert oder Median der zugehörigen Tagesganglinien) und deren Werte direkt zur Prognose genutzt.

Rein kalenderbasierte Prognosen können aktuelle Verkehrssituationen naturgemäß nicht berücksichtigen und werden daher typischerweise für größere Prognosehorizonte eingesetzt (ab etwa einer Stunde, abhängig von Netz- und Nachfragecharakteristik). Für diese Prognosehorizonte hat die aktuelle Verkehrssituation wenig bis keinen beschreibenden Charakter. Daher sind die tagesstypischen Werte in der Regel die bestmögliche Schätzung. Kalenderbasierte Prognosen können außerdem als Rückfallebene im Rahmen komplexerer Systeme bei fehlenden aktuellen Daten verwendet werden. Schließlich sollte die kalenderbasierte Prognose bei der Entwicklung komplexerer Prognoseverfahren unter Einbindung von Verkehrsdaten als Referenzverfahren betrachtet werden.

Werden die Verkehrsdaten vor der Mustererkennung klassifiziert, so sollte eine Klassenhierarchie definiert werden, an deren aggregiertem Ende wieder die gesamte Datenbasis steht. So könnte ein Verkehrsmuster auch in einer anderen Klasse gefunden werden, wenn es deutlich ähnlicher als das ähnlichste der eigenen Klasse wäre.

3.2.2 Segmentierung von Verkehrsmustern

Verkehrsmuster, z.B. Tagesganglinien, weisen bei Beobachtung über Monate und Jahre wiederkehrende Eigenschaften auf. Die Idee der Segmentierung ist das Bilden von repräsentativen Mustern für typische Situationen mit dem Ziel, fortan diese statt der zugrunde liegenden Muster zu nutzen.

Es existieren verschiedene Methoden zur optimalen Segmentierung einer Datenbasis. Grundsätzlich kann zwischen hierarchischen und partitionierenden Verfahren unterschieden werden. Einsatzmöglichkeiten der Clusteranalyse zur Segmentierung von Verkehrsdaten

werden in verschiedenen Arbeiten beschrieben. KLEINHANS [2002] beschreibt beispielsweise verschiedene Clusterverfahren zur Bildung von Ganglinienprototypen für Autobahn-Messstellen in Nordrhein-Westfalen. Eine Anwendung ist die Bildung von "Risikoganglinien" zur statistisch gesicherten Aussage über maximale Staulängen. Im ATHENA System wird eine tagesunabhängige Segmentierung zur Bildung von Prototypen vorgenommen, d.h. die Cluster sind unabhängig von eventuell vorhandenem Vorwissen gebildet worden [DANECH-PAJOUH UND ARON 1991].

Die Segmentierung von Verkehrsmustern bietet den Vorteil, dass die Datenbasis reduziert wird und gleichzeitig typische Muster betont werden können. Je nach Applikation kann es sich nachteilig auswirken, dass ein Teil der Information verloren geht. Bei nicht optimaler Wahl der Parameter können wesentliche Muster verloren gehen. Darüber hinaus kann es zur Prognose relevant sein, ob ein bestimmtes Muster schon mehrfach beobachtet wurde, oder ob es sich um eine Einzelercheinung handelt.

3.3 Regressionsanalytische und zeitreihenbasierte Verfahren

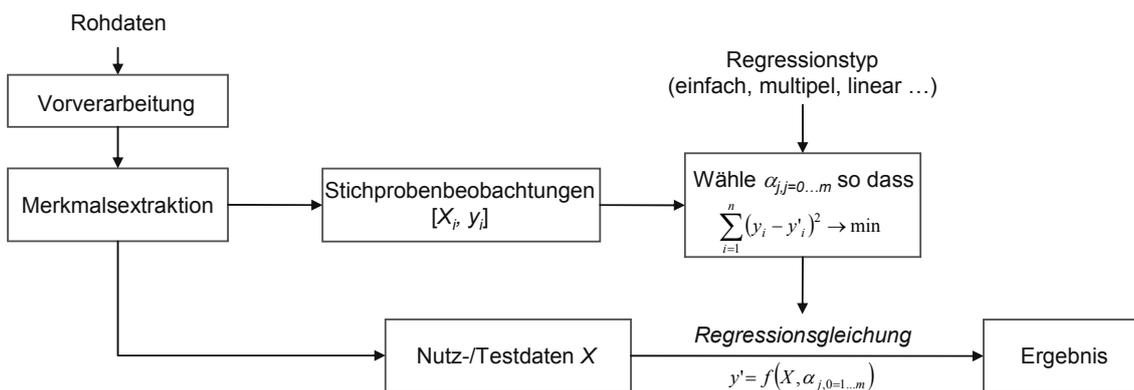


Abb. 3.3 Ablauf der Regressionsmodellbildung

Mit der Regressionsanalyse wird versucht, einen funktionalen Zusammenhang zwischen einer (*einfache Regression*) bzw. mehreren (*multiple Regression*) unabhängigen Variablen x_j und einer abhängigen Variablen y zu finden und diesen geschlossen funktional zu formulieren. Allgemein hat ein Regressionsanalytisches Modell oder Regressionsmodell die Form:

$$y'_i = f(X_i, \alpha_j) \quad i = 1, \dots, n; \quad j = 0, \dots, m; \quad X_i = [x_{i,1}, \dots, x_{i,j}, \dots, x_{i,n}]$$

Im Fall der *linearen* Regression lautet die Regressionsformel:

$$y'_i = \alpha_0 + \sum_{j=1}^m \alpha_j \cdot x_{i,j}$$

Die Parameter α_j werden meist nach der Methode der kleinsten Quadrate geschätzt [SACHS 2000 S. 489ff]. Ermittelt werden also Parameter in der Art, dass – bezogen auf die zugrunde liegende Trainingsdatenmenge $\{y_i, X_i\}$ – die Summe der quadrierten Abweichungen zwischen Schätzung und Beobachtung minimal wird.

$$\sum_{i=1}^n \left[y_i - \alpha_0 + \sum_{j=1}^m \alpha_j \cdot x_{i,j} \right]^2 \rightarrow \min$$

Bei der *nicht linearen Regression* wird statt des linearen Zusammenhanges ein anderer Funktionstyp (z.B. Polynom, Exponentialfunktion) vorgegeben und an die Trainingsdaten angepasst.

RICE UND ZWET [2001] verwenden ein einfaches Regressionsmodell zur Kurzfristprognose von Reisezeiten auf Schnellstraßen auf Basis von lokalen Messungen. Die Parameter werden zeitabhängig auf Basis aktueller und nahvergänger Beobachtungen geschätzt. YANG ET AL. [2000] verwenden ein lineares Regressionsmodell mit zwei unabhängigen Variablen (Verkehrsstärke, Belegung) zur Prognose der Reisezeit auf einem Schnellstraßenabschnitt.

Einen Spezialfall der regressionsanalytischen Modelle sind die Verfahren der Zeitreihenanalyse. Diese versuchen, eine beobachtete Zeitreihe aus ihrem bisherigen Verlauf fortzuschreiben. Vertreter dieser Verfahren sind beispielsweise die Trendextrapolationen, exponentielle Glättungsmodelle, sowie die von BOX UND JENKINS [1970] begründeten ARMA-Modelle und deren Varianten (ARMAX, ARIMA).

ARMA (Auto Regressive Moving Average) Modelle und deren Varianten besitzen zwei Grundmodelle. Das AR(p) Modell erklärt eine Variable x zum Zeitpunkt t mit einem multiplen Regressionsmodell aus seinen p Vorgängerwerten:

$$x(t)_{AR} = \alpha_1 \cdot x(t-1) + \alpha_2 \cdot x(t-2) + \dots + \alpha_p \cdot x(t-p) + \varepsilon(t)$$

Dabei ist der Rauschterm ε eine Zufallszahl aus einer Normalverteilung mit Erwartungswert Null und die Parameter α werden durch Fehlerminimierung geschätzt. Das MA(q) Modell erklärt x zum Zeitpunkt t durch die q vorangegangenen Prognosefehler ε :

$$x(t)_{MA} = \beta_1 \cdot \varepsilon(t-1) + \beta_2 \cdot \varepsilon(t-2) + \dots + \beta_q \cdot \varepsilon(t-q)$$

Dabei werden die Parameter β durch Fehlerminimierung geschätzt. ARMA Modelle und deren Derivate finden häufig Anwendung für Prognosen, insbesondere für kurze Prognosehorizonte ("One-Step-Ahead" Prognosen). BELZNER [2002] setzt ein univariates ARIMA Modell als Teil eines hybriden Verfahrens zur Prognose der lokalen Geschwindigkeit auf Autobahnen ein. Aufgrund seiner Fähigkeit, lineare Abhängigkeiten in Zeitreihen gut abzubilden, wird das ARIMA Modell dabei zur Prognose im freien Verkehr eingesetzt.

Der Nachteil von ARMA Modellen ist, dass die zu prognostizierende Zeitreihe stationär sein muss. Dadurch wird die Prognose einer in der Zeitreihe nicht erkennbaren Änderung (z.B. Stau) aus sich selbst im Prinzip unmöglich. Gerade diese Eigenschaft kann wiederum zur Erkennung von Störungen genutzt werden, in dem die Abweichung des prognostizierten Wertes der Zeitreihe von der Realität abweicht [FERRARI 1988].

Ein um eine räumliche Komponente erweitertes Zeitreihenmodell stellen KAMARIANAKIS UND PRASTACOS [2002] mit ihrem multivariaten STARIMA (Space Time Autoregressive Integrated Moving Average) Modell vor. Dabei werden Verkehrsstärkezeitreihen (aggregiert auf 7,5 Minuten) von 25 sich im Netz stromaufwärts befindlichen Messstellen zur Kurzfristprognose der Verkehrsstärke an einer Messstelle (Abb. 3.4) genutzt.

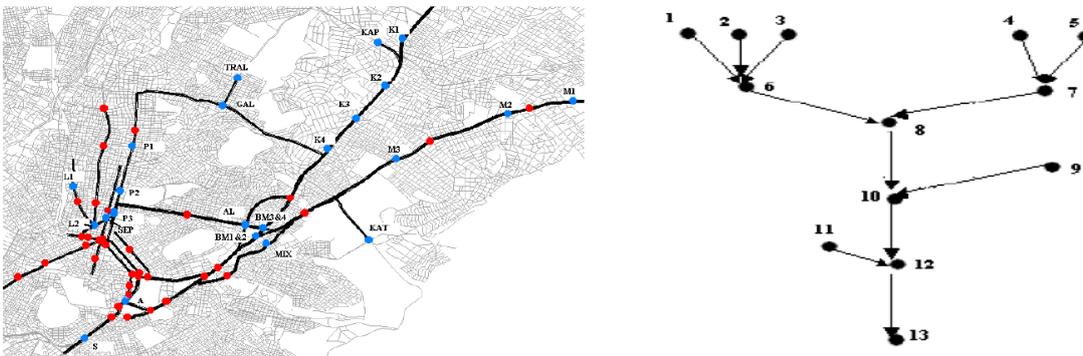


Abb. 3.4 Zur Prognose verwendete Messstellen (links) und deren Repräsentation als Baum (rechts) [KAMARIANAKIS UND PRASTACOS 2002]

Die prognostizierte Verkehrsstärke wird als gewichtete Linearkombination der Messwerte der stromaufwärts gelegenen Messstellen ermittelt, wobei die Gewichte zeitlich und entsprechend des Abstandes der jeweiligen Messstelle berechnet werden. Die Modellparameter wurden für unterschiedliche Tageszeiten angepasst.

3.4 Künstliche Neuronale Netze

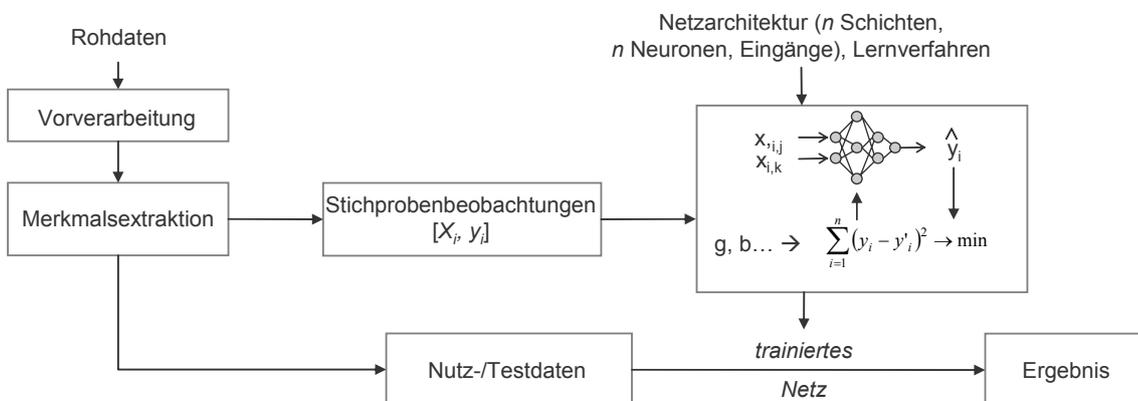


Abb. 3.5 Mustererkennung mit künstlichen Neuronalen Netzen

Künstliche Neuronale Netze (engl.: Artificial Neural Networks *ANN*) sind Funktionsapproximatoren, deren Struktur an die des menschlichen Gehirns angelehnt ist. Es existieren viele verschiedene Netztopologien, wobei in der Mustererkennung die mehrschichtigen vorwärtsgerichteten Netze (engl.: *Multi Layer Feedforward MLF* oder auch *Multi Layer Perceptron MLP*) die wohl am häufigsten eingesetzte Klasse ist.

Ein *MLF* besteht aus einer Eingabeschicht (die den Eingabevektor X entgegennimmt), einer Ausgabeschicht (welche die abhängige Variable y ausgibt) und einer oder mehreren versteckten Schichten. Die Schichten bestehen aus Neuronen, welche miteinander verbunden sind. Neuronen transformieren ein Eingangssignal durch Übertragungs-/Aktivierungsfunktion, einen Gewichtungsfaktor und einen konstanten Bias. Das Eingangssignal für die Neuronen bestimmt sich wiederum aus den Ausgangssignalen der vorgeschalteten Neuronen. *MLF* lernen den funktionalen Zusammenhang zwischen Eingabe und Ausgabe auf Basis von Trainingsmustern. Dabei werden die Gewichte iterativ in Richtung des minimalen Fehlers verändert (Gradientenabstiegsverfahren, meist Backpropagation und verschiedene Derivate). Vorteile von *MLF* und *ANN* sind, dass sie im Prinzip jede Funktion (so weit vorhanden) nachbilden können und dass effiziente Lernalgorithmen zur Netzanpassung zur Verfügung stehen. Nachteilig ist, dass es sich bei einem *ANN* um eine "Black-Box" handelt und der funktionale Zusammenhang nicht einfach erkennbar ist und es folglich schwierig ist, ein Ergebnis nachzuvollziehen. Dies kann insbesondere bei unbekanntem Eingabemustern ein Problem darstellen. Ferner können *ANN* je nach Entwurf, Aufgabenstellung und Trainingsdatensatz zu einer Überanpassung an den Trainingsdatensatz und somit zu einer fehlenden Generalisierungsfähigkeit tendieren.

In Prognoseanwendungen erfreuen sich *ANN* großer Beliebtheit. So verwendet VON DER RUHREN [2006] ein *MLF* zur Klassifizierung lokaler, messstellenbezogener Verkehrssituationen zu einer auf die Messstellengruppe bezogene Verkehrslage. Häufig werden Künstliche Neuronale Netze zur Kurzfristprognose lokaler Verkehrskenngrößen auf Schnellstraßen eingesetzt. BELZNER [2002] verwendet ein *MLF* zur Prognose der lokalen Geschwindigkeit in gestauten Verkehrszuständen unter Einbezug von Messwerten stromabwärtiger ("Rückstaukomponente") und stromaufwärtiger ("Nachfragekomponente") Messquerschnitte. TAO ET AL. [2006] vergleichen drei Netzarchitekturen (ein *MLF*, ein modulares Netz durch Zusammenschluss mehrerer einfacher *MLF* und ein *MLF* mit vorgeschalteter Hauptkomponentenanalyse) zur Prognose von Reisezeiten auf Schnellstraßen. Dabei hat sich kein Netz als in allen Situationen überlegen gezeigt. Es wird empfohlen, erweiterte Architekturen zur situationsbezogen optimalen Prognose zu erforschen. JIANG & ZHANG [2001] trainieren ein *MLF* zur Prognose von Reisezeiten auf Schnellstraßen ausschließlich mit Verkehrsstärken als Eingangsgrößen, was aufgrund der "Uneindeutigkeit" der Verkehrsstärke zur Charakterisierung des Verkehrszustandes eher ungewöhnlich ist. Im Gegensatz zu *MLF* berücksichtigen Rekurrente Neuronale Netze (*RNN*) den zeitlichen Verlauf der Eingangsgrößen durch die interne Netzstruktur und nicht über eine erweiterte Eingabeschicht. Dies kann bei der zeitreihenba-

sierten Prognose (z.B. Reisezeitprognose auf Schnellstraßen [VAN LINT 2004], Stauprognose auf Schnellstraßen [ZHOU UND NELSON 2002]) Vorteile bringen.

3.5 Selbstlernende Expertensysteme

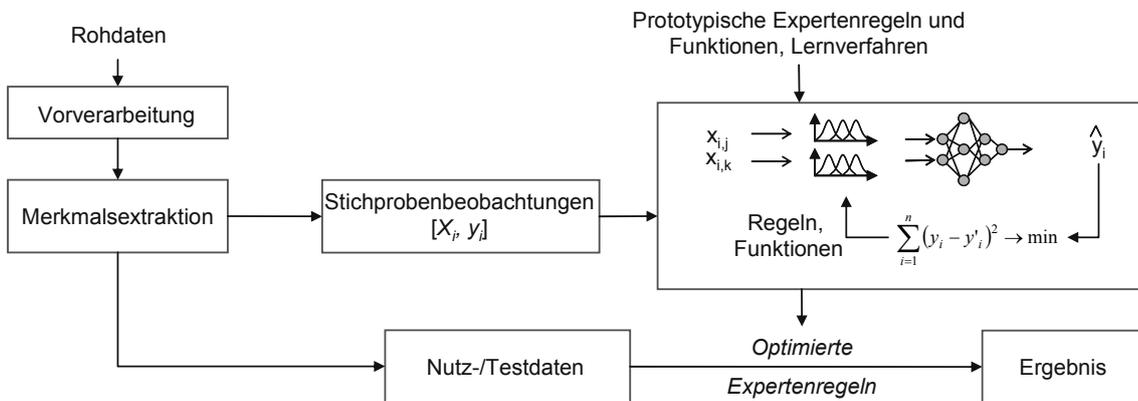


Abb. 3.6 Selbstlernendes Expertensystem

Expertensysteme sind allgemein Systeme, die auf Basis von Expertenwissen für eine gegebene Ausgangssituation eine Lösung ermitteln. Es existieren unterschiedliche Modelle zur Repräsentation und Nutzung des Expertenwissens. Neben fallbasierten Systemen und Entscheidungsbäumen stellen regelbasierte Systeme eine wichtige Klasse dar. Regelbasierte Expertensysteme unterscheiden sich von rein datenorientierten Verfahren dadurch, dass vorhandenes Wissen in Form von formulierten Regeln genutzt wird, z.B.:

WENN Verkehrsstärke IST *niedrig* UND Belegung IST *sehr hoch* DANN Reisezeit IST *hoch*

Fuzzy-Expertensysteme erlauben dabei auch eine unscharfe Formulierung des Wissens. Bei der Abbildung komplexer Prozesse wie der Prognose verkehrlicher Kenngrößen, ist es meist notwendig, Expertensysteme zusätzlich anhand von Eingabe-/Ausgabemustern zu trainieren.

Häufig werden dazu Neuro-Fuzzy Systeme (NFS) eingesetzt. Der Vorteil von NFS ist die Kombination aus explizit formulierbaren qualitativen Regeln der Fuzzy Logik und die Anpassung der formulierten Zugehörigkeitsfunktions-Prototypen an die vorhandenen Beobachtungen mit Trainingsmethoden aus dem Bereich der künstlichen Neuronen Netze.

LINAUER ET AL. [2006] setzen ein ANFIS (Adaptive Neuro-Fuzzy Inference System) zur Schätzung von Routengeschwindigkeiten ein. Die Länge der betrachteten Routen ist zwischen 0,9 und 2,9 km. Als Eingangsgröße werden auf der jeweiligen Route gemessene lokale Geschwindigkeiten und Verkehrsstärken verwendet, die Ausgabe-Trainingsdaten zur Optimie-

rung des Fuzzy Systems werden aus fahrzeuggenerierten Daten (aus Taxipositionsmeldungen) ermittelt.

Im Gegensatz zu einem "normalen" FIS, in dem die Regeln manuell erstellt und die Zugehörigkeitsfunktionen manuell kalibriert werden müssen, wird diese Aufgabe von einem Neuronalen Netz übernommen. Trainingsziel ist die Minimierung des mittleren Schätzfehlers. Der gelernte Zusammenhang zwischen lokaler Geschwindigkeit, Verkehrsstärke und mittlerer Routengeschwindigkeit ist in Abb. 3.7 beispielhaft dargestellt.

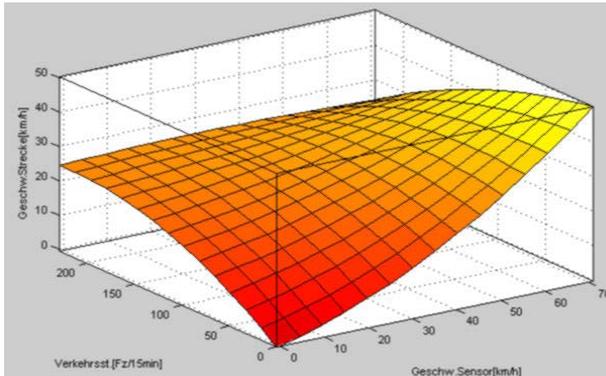


Abb. 3.7 Gelernter Zusammenhang zwischen v_{lokal} , q und v_{Route} [LINAUER ET AL. 2006]

Zur weiteren Verarbeitung werden die geschätzten Routengeschwindigkeiten mit aktuell ermittelten Routengeschwindigkeiten fusioniert, so diese vorhanden sind.

3.6 Instanzbasiertes Lernen

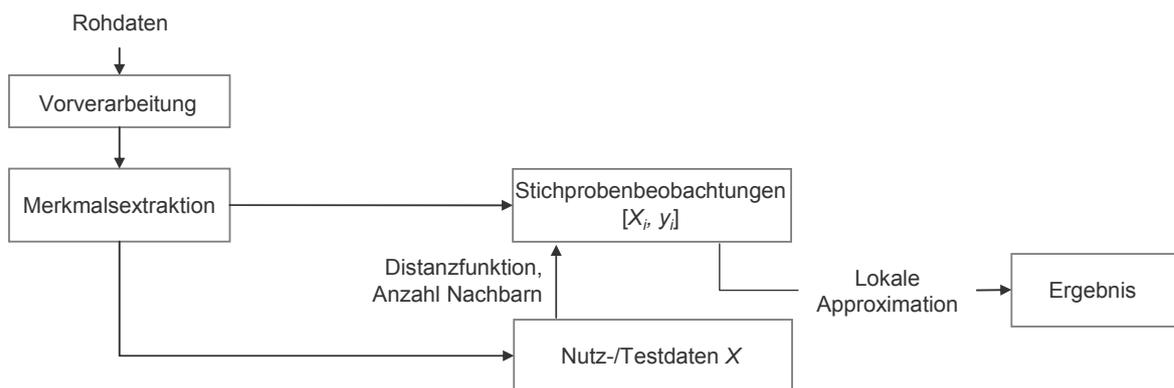


Abb. 3.8 Instanzbasiertes Lernen

Instanzbasierte Lernverfahren (auch: nichtparametrische Regression) werden in Kapitel 4 detailliert dargestellt. Daher wird an dieser Stelle nur kurz das Prinzip erläutert. Anschließend werden verschiedene Anwendungsbeispiele beschrieben. Instanzbasiertes Lernen kann mit einem Zitat von MOORE [2001] treffend beschrieben werden:

"To make a prediction, search database for similar datapoints, and fit with the local points."

Beim Instanzbasierten Lernen wird also keine globale Funktion zur Abbildung des gesamten Wertebereichs gebildet, sondern die vermutete Funktion durch lokale Teillösungen direkt aus den Beobachtungen angenähert.

Zunächst wird die Ähnlichkeit des aktuellen Eingangsmusters zu den historischen Mustern ("Instanzen") berechnet. Dazu muss ein sinnvolles Distanzmaß definiert werden. Die k ähnlichsten Instanzen (Nachbarn) werden zur Bildung der Prognose berücksichtigt. Dabei kann k zwischen 1 und der Anzahl der Instanzen variieren. Die Instanzen können bei der Prognosebildung mit einer Gewichtung versehen werden. Die Gewichtung wird normalerweise aus der Ähnlichkeit der Instanzen zum aktuellen Muster berechnet. Die lokale Approximation aus den " k Nächsten Nachbarn" kann mit verschiedenen Methoden, z.B. mit kernbasierter Dichteschätzung oder lokal gewichteter linearer Regression durchgeführt werden.

WILD [1996] stellt ein Verfahren zur Kurzfristprognose von Verkehrsstärken auf Basis historischer Verkehrsstärkeganglinien vor. Dabei werden historische Ganglinien geglättet und transformiert (siehe Abb. 3.9), klassifiziert und in Repräsentanten segmentiert. Zur Prognose wurde der zur aktuellen Ganglinie entsprechend dem vertikalen Abstand an den Extremstellen ähnlichste Repräsentant der gültigen Tagesklasse verwendet. Dabei wurde jede Messstelle isoliert betrachtet. Der eigentliche Prognosewert wurde aus dem Wert des Repräsentanten zum prognostizierenden Zeitpunkt sowie der Differenz zwischen dem aktuell gemessenen Wert und dem entsprechenden Wert des Repräsentanten berechnet. Das Verfahren lieferte für einen Prognosehorizont von 9 Minuten bessere Ergebnisse als die Trivialprognose und ein als einfaches, auf einem gleitenden Mittelwert basierendes Prognoseverfahren. Bei der Optimierung des Verfahrens stellte sich heraus, dass sich eine größere Anzahl an Repräsentanten günstig auf die Prognosegüte auswirkt.

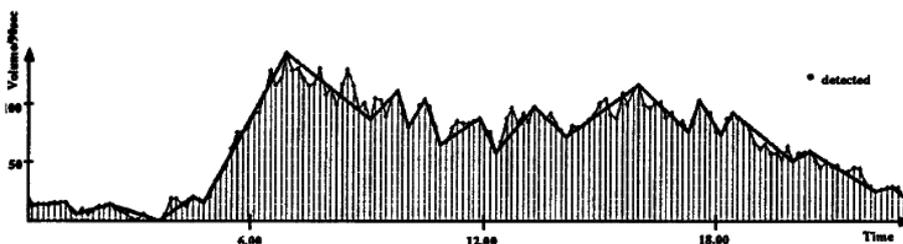


Abb. 3.9 Verkehrsstärke und transformierte Ganglinie [WILD 1996]

VON DER RUHREN [2006] beschreibt ein zweistufiges Mustererkennungsverfahren zur Prognose von Verkehrsstärken. In der ersten Stufe ("lokale Ebene") wird für jeden Messquerschnitt im betrachteten Netz die wahrscheinlichste Tagesganglinie aus einer Wissensbasis ausgewählt. Dabei liegen so genannte Standardklassen (Wochentage) und Sonderklassen (verkehrswirksame Ereignisse) vor. Zur Auswahl der wahrscheinlichsten Tagesganglinie wird ein neu entwickeltes Verfahren zur Beurteilung der Ganglinienähnlichkeit eingesetzt. Die Ähn-

lichkeit wird über den Abstand der einzelnen Stützstellen der letzten n Intervalle berechnet. Im kritischen Bereich (also bei hohen Verkehrsstärken) wird generell ein geringerer Abstand zugelassen. Dadurch wird gewährleistet, dass die relevanteren Zustandsbereiche stärker gewichtet werden und Abweichungen z.B. in Schwachlastzeiten nicht überbewertet werden.

FÖRSTER [2008] stellt ein Verfahren vor, in dem die Verkehrsstärken für 52 Messstellen durch einen Abgleich der aktuellen Messungen mit einer historischen Datenbasis für einen Prognosehorizont von 20 Minuten geschätzt werden. Die historische Datenbasis besteht aus auf 20 Minuten aggregierten Verkehrsstärkewerten für sämtliche Messstellen aus einem Zeitraum von etwa fünf Monaten. Bei der Prognose wird das zu den aktuellen Messungen ähnlichste Muster in einer nach Tagesklassen sortierten Datenbasis gesucht. Zur Berechnung der Ähnlichkeit wird ein Raum-Zeitfenster über alle Messstellen aufgespannt und Intervall für Intervall über die historische Datenbasis "geschoben" (siehe Abb. 3.10).

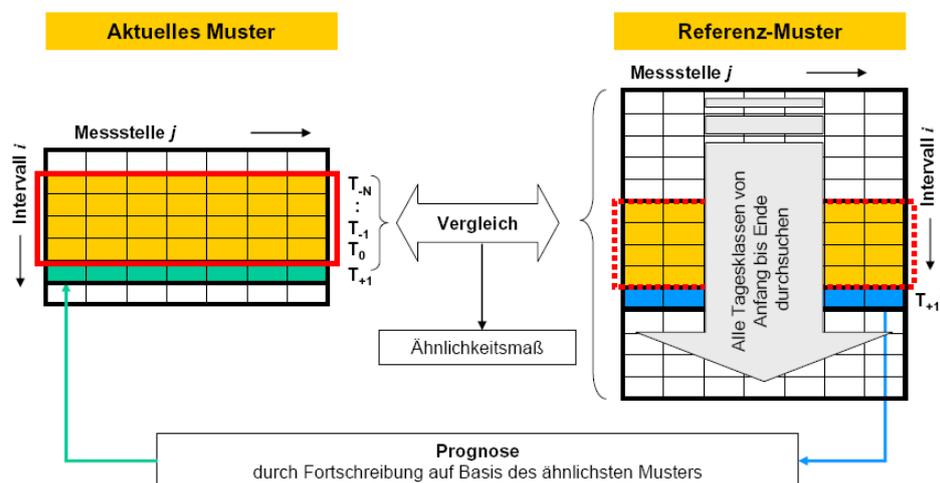


Abb. 3.10 Prinzipskizze des Prognoseverfahrens von FÖRSTER [2008]

Für jede Position des Fensters wird ein Ähnlichkeitsmaß berechnet. Es wurden zwei Distanzmaße zur Berechnung der Ähnlichkeit verwendet: (1) Die mittlere quadratische Abweichung zwischen den beiden Mustern und (2) ein Korrelationsmaß. Die prognostizierte Verkehrsstärke für die Messstellen wird dann aus dem ähnlichsten Muster abgeleitet. Zur Anpassung der prognostizierten Werte an den aktuellen Messwert wurden zwei Methoden untersucht: (1) Skalierung durch einen Faktor der die absolute Abweichung zwischen dem aktuellen und dem Referenzmuster berücksichtigt und (2) Linearkombinationen der Messwerte in den vorgelagerten Intervallen. Insgesamt wurden die besten Ergebnisse durch Verwendung des Korrelationsmaßes und der Skalierung der Prognosewerte über die absolute Abweichung zwischen aktuellem und Referenzmuster erreicht. Mit diesem vergleichsweise einfachen Ansatz konnten relativ gute Ergebnisse erzielt werden, die von komplexeren Verfahren (z.B. mit auf Basis von Genetischen Algorithmen optimierten Neuronalen Netzen) nur geringfügig übertroffen werden.

Im Rahmen des Forschungsprojektes iQmobility wurde für Berlin ein Verfahren zur Kurzfrist-Prognose der innerörtlichen zeitlich-räumlichen Störungsausbreitung auf Basis räumlicher Mustererkennung entwickelt [MAIER ET AL. 2007, MAIER ET AL. 2008]. Diese basiert auf Verkehrszuständen, die für jedes Zeitintervall auf jeder Kante des Berliner Straßennetzes durch ein dynamisches Routenwahl- und Umlegungsmodell (MONET/VISUM-online) berechnet werden. Aus umfangreichen historischen Verkehrslagedaten wurden Störungsmuster generiert. Zur Prognose wird das aktuelle Verkehrsmuster ("Schnappschuss") auf die historische Datenbasis gematcht um ähnliche Muster zu finden. Die zeitliche Störungsentwicklung wird entsprechend der historisch-zeitlichen Entwicklung des jeweils gefundenen Referenzmusters für einen Prognosehorizont von bis zu 60 Minuten prognostiziert.

ROBINSON [2005] stellt ein Verfahren zur Schätzung von Reisezeiten vor. Dabei werden Reisezeiten der Kennzeichenerfassung aus dem Londoner City-Maut System als abhängige Variable und lokale Detektordaten q und b aus dem SCOOT (Split Cycle Offset Optimisation Technique) System als unabhängige Variable betrachtet. Die betrachteten Detektoren lagen jeweils auf der betrachteten Route. Das Verfahren wurde auf zwei Routen getestet (~1km mit 3 Messstellen auf der Route und ~0,7 km mit 4 Messstellen auf der Route, einer direkt stromaufwärts und einer direkt stromabwärts). Das entwickelte Verfahren lieferte bessere Ergebnisse als die zum Vergleich hinzugezogenen Verfahren (Naive Schätzung, Kalenderbasiert, einfache Regressionsanalyse, künstliches Neuronales Netz). Weitere Erkenntnisse waren, dass eine Parameteranpassung für unterschiedliche Stecken nicht notwendig ist (Übertragbarkeit). Auch wurde eine ausführliche Parameterbetrachtung und Optimierung durchgeführt. ROBINSON [2005] stellt in einem Ausblick das Potential der Reisezeitschätzung basierend auf historischen mit FCD ermittelten Reisezeiten und zugehörigen lokalen Messungen heraus.

TUROCHY UND PIERCE [2004] untersuchten *k-Nächste-Nachbarn* Verfahren zur Prognose der lokalen Geschwindigkeit auf einer Schnellstraße für ein Zeitintervall. Als Merkmalsvektor wurde die Geschwindigkeit der jeweiligen Messstelle der letzten drei Zeitintervalle genutzt und mit einer nach Wochentagen klassifizierten Datenbasis verglichen. Es wird dabei eine Gewichtung der Instanzen entsprechend der Ähnlichkeit als wichtige Eigenschaft identifiziert. DAVIS UND NIHAN [1991] entwickelten ein Verfahren zur Prognose von lokalen Verkehrsgrößen q und b auf einer Schnellstraße. Eingangsgrößen waren q und b aus dem vorhergehenden Intervall von mehreren Messstellen. Bei den Untersuchungen wurde die Berücksichtigung nur eines Nachbarn k als optimal herausgefunden. SMITH UND DEMETSKY [1997] fanden heraus, dass ihr Instanzbasiertes Lernverfahren zur Prognose der Verkehrsstärke bezüglich Genauigkeit und Übertragbarkeit den zum Vergleich hinzugezogenen ARIMA Modellen und einem künstlichen Neuronales Netz überlegen war. CLARK [2003] stellt eine Methode zur Prognose von q , v und b auf Basis von 10-Minuten Daten auf Schnellstraßen vor. Dabei beinhaltet eine Instanz (Muster) jeweils ein Tripel $\{q,v,b\}$ zur Berücksichtigung der "3-dimensionalen" Eigenschaft des Verkehrszustands. CHROBOK [2005] entwickelt und testet

einen Prognosealgorithmus auf Basis von lokalen Verkehrsstärkedaten aus dem nordrhein-westfälischen Autobahnnetz. Dazu wird eine Nächste-Nachbarn-Suche auf historischen, segmentierten Ganglinien durchgeführt. Zur Prognose werden die aktuellen Messwerte in Abhängigkeit des Prognosehorizontes eingebunden. OSWALD ET AL. [2000] beschäftigten sich mit der Parametersensitivität und der Effizienzoptimierung durch Strukturierung und Indizierung von hochdimensionalen Datenbasen. HANDLEY ET AL. [1998] schätzten Reisezeiten aus q und b . Dabei fanden sie heraus, dass die Verkehrsvariablen einen stärkeren Einfluss haben als Kalendervariable wie Tageszeit und Wochentag. YOU UND KIM [2000] nutzten fahrzeuggenerierte Daten der vorhergehenden Stunde zur Prognose von Reisezeiten auf Basis eines k Nächste Nachbarn Algorithmus innerorts.

3.7 Optimierung der Modellparameter

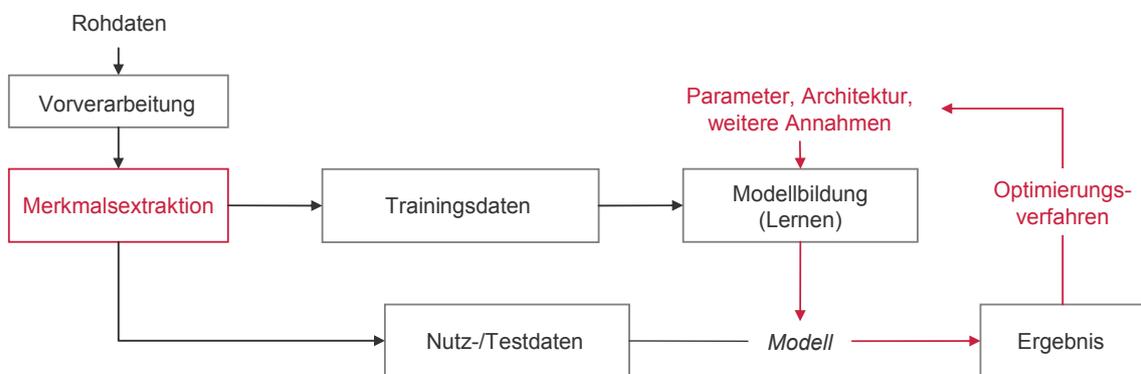


Abb. 3.11 Optimierung von Verfahren zur Mustererkennung

Während die optimale Anpassung an die Trainingsdaten ein Teil des Mustererkennungsprozesses und damit automatisiert ist, bleibt der Modellentwurf dem Entwickler überlassen. Neben der Merkmalsselektion müssen je nach Modellklasse weitere Entwurfsparameter festgelegt werden (siehe Tab. 3.1).

Modell	Zu spezifizierende Entwurfsparameter
Künstliche Neuronale Netze	Auswahl der geeigneten Netztopologie (Anzahl Input-Neuronen, Anzahl versteckter Schichten), Verknüpfungen, Bias.
Regressionsanalytische Modelle	Spezifikation der Form (linear, nicht linear, ...).
Instanzbasiertes Lernen	Distanzmetrik, Anzahl zu berücksichtigender Nachbarn, Parameter der lokalen Approximationsfunktion.
Selbstlernende Expertensysteme	Prototypische Regeln/Fälle, Struktur des Lernverfahrens

Tab. 3.1 Modellklassen und zugehörige Entwurfsparameter

Grundlage ist ein sinnvoller Entwurf auf Basis der Erfahrung des Entwicklers und unter Einbezug von möglicherweise verfügbarem Expertenwissen. Mittels einer Sensitivitätsanalyse sollte anschließend die Prognosequalität für Modellparametervarianten untersucht werden.

Nach einer sinnvollen Modellwahl und der Identifizierung der wichtigsten Entwurfsparameter können diese auf Basis von Trainingsmustern manuell oder mit Optimierungsverfahren automatisch eingestellt werden. Für eine automatische Optimierung lässt sich das Problem üblicherweise nicht als ableitbare Funktion formulieren, so dass gradientenbasierte Verfahren nicht eingesetzt werden können.

Zur Optimierung eines Parameters (Anzahl der zu berücksichtigenden Nachbarn k) nutzt ROBINSON [2005] eine Fibonacci-Suche (Fibonacci line search). Die Fibonacci-Suche ist eine Methode der eindimensionalen Optimierung ohne Verwendung von Ableitungen. Zur Optimierung mehrerer Parameter werden häufig stochastische/heuristische Optimierungsverfahren wie simuliertes Abkühlen (*engl.: Simulated Annealing*), Schwarmoptimierung oder Genetische Algorithmen (GA) eingesetzt. Anwendungen finden sich beispielsweise im Bereich der Finanzmathematik (NATH ET AL. [2005] schlagen einen GA zur Auswahl der wichtigsten Merkmale, wie Einkommen, Vermögen und Berufserfahrung, zur Prognose der Kreditwürdigkeit vor) und der Medizin (JAIN UND MAZUMDAR [2003] beschreiben ein hybrides Verfahren zur Diagnose von Brustkrebs, wobei die k ähnlichsten Trainingsmuster zur Erkennung genommen und mittels eines GA eine Clusterung auf wesentliche Referenzmuster durchgeführt wird).

BAJWA ET AL. [2003] verwenden einen GA zur Optimierung der Parameter eines Instanzbasierten Lernverfahrens zur Reisezeitprognose auf einem Autobahnabschnitt bei Tokio/Japan. Dabei wurden optimiert:

- Anzahl der zu berücksichtigenden Messstellen (alle berücksichtigten Messstellen liegen auf der Route, für die die Reisezeit prognostiziert werden soll).
- Anzahl der zu berücksichtigenden Nachbarn k .
- Zeitliche Ausdehnung des Musters (wobei das Gewicht der Merkmale linear mit dem Alter abnimmt). Die mögliche Ausdehnung wurde dabei mit 10 bis 160 Minuten vorgegeben.
- Betrachtetes Zeitfenster. Es werden nur Instanzen auf Ähnlichkeit untersucht, die bezogen auf die Tageszeit um nicht mehr als t_{max} von der aktuellen Tageszeit abweichen. t_{max} wird zwischen ± 15 und ± 120 Minuten variiert.

Erkenntnisse waren: (1) die optimalen Parametereinstellungen sind nicht global gültig sondern zeitabhängig, (2) die optimale zeitliche Ausdehnung ist mit typischerweise 10 Minuten relativ gering, (3) die optimale Größe des betrachteten Zeitfensters variiert deutlich zwischen 15 und 120 Minuten, (4) das optimale k variiert zwischen 5 und 40. Die Vermutung, dass eine mit dem Datenalter abnehmende Gewichtung sinnvoll ist, hat sich bestätigt. Die Optimierung wurde nicht für unterschiedliche Prognosehorizonte durchgeführt.

Heuristische Verfahren, insbesondere Genetische Algorithmen, scheinen aus mehreren Gründen für die Optimierung von Parametern in Mustererkennungsverfahren gut geeignet zu sein. Zum einen sind sie in der Lage, die Modellparameter gleichzeitig und ohne Formulierung einer geschlossenen lösbaren Zielfunktion zu optimieren. Zum anderen wurde die erfolgreiche Anwendung schon in vielen Fällen gezeigt.

3.8 Modellauswahl und weiteres Vorgehen

Verfahren der Mustererkennung werden für verschiedene Schätz- und Prognoseaufgaben im Verkehrsmanagement eingesetzt, so dass ihr Einsatz für die hier beabsichtigte Prognose von Reisezeiten und lokalen Verkehrskenngrößen aus unvollständigen Beobachtungen prinzipiell gerechtfertigt erscheint.

Ein bewertender Vergleich der unterschiedlichen Ansätze aufgrund der in der Literatur vorgestellten Ergebnisse ist schwierig, da die Implementierung der Verfahren teilweise nicht detailliert beschrieben ist und die Randbedingungen in den Untersuchungen variieren. Zur begründeten Auswahl werden die verschiedenen Modelle im Folgenden bezüglich der Erfüllung der in Abschnitt 1.2 definierten Anforderungen beurteilt.

Klassifikations- und Segmentationsverfahren stellen keine Mustererkennungsverfahren im eigentlichen Sinne dar, sondern können beispielsweise zur Vorverarbeitung der Daten eingesetzt werden. Regressionsanalytische Modelle werden aufgrund der zu erwartenden großen Zahl von unabhängigen Variablen und der unbekannt, vermutlich stark nichtlinearen Funktionsform nicht näher in Betracht gezogen, da die so entstehenden Modelle komplex sind und eine große Zahl an Parametern besitzen (siehe beispielsweise SARLE [1994]). Insbesondere ist der Parametrierungsaufwand und der Aufwand zum Herstellen der Adaptivität als relativ hoch zu bewerten. Die Forderung der Robustheit gegenüber fehlenden Messwerten ist ebenfalls nur durch komplex zusammengestellte Regressionsgleichungen zu erfüllen, da bei geschlossenen Regressionsgleichungen nicht ohne weiteres Eingangswerte entfallen dürfen. Zeitreihenbasierte Prognosemodelle gehen von äquidistanten und vor allem vollständig beobachteten Zeitreihen aus. Diese Voraussetzung ist bei der gegebenen Datenverfügbarkeit mit zufällig beobachteten Verkehrskenngrößen im Allgemeinen nicht erfüllt. Künstliche Neuronale Netze sind zur Abbildung der vermuteten Zusammenhänge prinzipiell gut geeignet. Sie haben jedoch den Nachteil, dass die trainierten Netze nicht mehr interpretierbar sind und die Ergebnisse somit schwer nachvollzogen werden können. Des Weiteren ist eine explizite Modellierung der Zuverlässigkeit zwar prinzipiell möglich (siehe beispielsweise VAN LINT 2004), jedoch sehr aufwändig. Expertensysteme erlauben den expliziten Einbezug von a priori Wissen und dessen Formulierung in menschenverständlicher Form. Die zur Anpassung an Beobachtungen notwendigen Trainingsmethoden sind allerdings sehr rechenintensiv und steigen mit der Anzahl der Parameter und der Größe des Eingangsvektors extrem an [HUISKEN 2006]. Damit sind die Modellbildung und die Herstellung der Adap-

tivität zumindest mit Schwierigkeiten behaftet. Instanzbasierte Lernverfahren bieten den Vorteil, dass sie intuitiv verständlich sind, dass das gesammelte Wissen offen und nicht als "Black Box" vorliegt und dass durch Hinzufügen von Instanzen in die Musterdatenbank eine natürliche Adaptivität gegeben ist. Außerdem muss aufgrund einer lokalen Approximation keine globale Funktion formuliert und gelernt werden, was bei zu erwartenden stark nicht-linearen Zusammenhängen günstig ist. Schließlich bieten Instanzbasierte Verfahren die Möglichkeit, statistische Kenngrößen zur Prognosezuverlässigkeit direkt aus den Instanzen abzuleiten. Hauptnachteil der Instanzbasierten Lernverfahren ist die benötigte Rechenzeit und der benötigte Speicherbedarf, da das Modell nicht in einer vorgeschalteten Offline-Trainingsphase generiert, sondern im Prinzip bei jeder Prognoseanforderung neu gebildet werden muss. Gegenüber diesem, bei der heute typischen Rechnerleistung lösbaaren Problem, steht der Vorteil, dass das Instanzbasierte Lernen die definierten Anforderungen erfüllt. Daher wird im folgenden Kapitel ein Instanzbasiertes Lernverfahren zur Prognose von Verkehrskenngrößen entwickelt.

4 Entwicklung eines neuen Prognoseverfahrens

"Wer in der Zukunft lesen will, muss in der Vergangenheit blättern."

André Malraux (1901-1976)

In *Kapitel 4* wird das in dieser Arbeit entwickelte Verfahren zur Prognose von Verkehrskenngrößen bezüglich seiner Eingangsgrößen, Verarbeitungsschritte und Parameter spezifiziert. Es werden mehrdimensionale Merkmalsvektoren aus verkehrlichen und externen Variablen definiert, anhand derer die Ähnlichkeit von zwei (Verkehrs-)Situationen bewertet werden kann. Anschließend werden Möglichkeiten zur Qualifizierung und Quantifizierung der Prognosezuverlässigkeit dargestellt und die Implementierung des Verfahrens beschrieben.

4.1 Grundlagen des Instanzbasierten Lernens

4.1.1 Die Idee

Instanzbasiertes Lernen (oder auch "*nicht parametrische Regression*") geht zur Approximation einer unbekanntem Funktion $y = f(X)$, $X = x_1, \dots, x_n$ davon aus, dass Datenpunkte bzw. Datenvektoren unabhängiger Variable lokal zur Approximation direkt auf den Trainingsdaten genutzt werden können. Einfach ausgedrückt bedeutet dies: Beim Instanzbasierten Lernen werden die Trainingsdatensätze direkt zur Prognose genutzt. Ein System zum Instanzbasierten Lernen wird durch vier Komponenten spezifiziert:

- Auswahl der Merkmale X
- Definition einer Distanzmetrik D
- Anzahl zu berücksichtigender Instanzen ("Nachbarn") k
- Lokale Approximation der Daten (bei Kernregression: Gewichtungsfunktion w)

Zur Verdeutlichung soll ein einfaches fiktives Beispiel dienen. Es soll die Verkehrsstärke q an einem Messquerschnitt zukünftig mittels Instanzbasierten Lernens geschätzt werden. Da für den Messquerschnitt historische Beobachtungspaare der Verkehrsstärke und der Geschwindigkeit vorliegen und angenommen wird, dass die Verkehrsstärke und die Geschwindigkeit einen Zusammenhang besitzen, soll der Merkmalsvektor X der unabhängigen Variablen aus der Geschwindigkeit v bestehen (1). Aus aktuellen Messungen v_{akt} soll nun das zu-

gehörige q'_{akt} geschätzt werden. Dazu werden die gespeicherten $v_{historisch}$ bezüglich ihrer Ähnlichkeit zu v_{akt} bewertet. Die Ähnlichkeit wird hier zum absoluten Abstand $D_i = |v_{akt} - v_{historisch, i}|$ für alle Instanzen i berechnet (2). Die Liste $[v_{historisch}, q_{historisch}]$ wird nach D aufsteigend sortiert. Zur Schätzung von q_{akt} werden die k ersten Listeneinträge $q_{historisch}$ ausgewählt (k Nächste Nachbarn). k wird hier zu 51 gewählt (3). q_{akt} wird im vorliegenden Fall als einfacher Mittelwert aus den ähnlichsten historischen Mustern lokal approximiert (4).

Abb. 4.1 zeigt die historischen Instanzen (blau) und die Schätzung von q aus v für $v = 50$ km/h aus den 51 Nachbarn (rot, links), sowie die entstehende Funktion für einen Wertebereich von 0 bis 70 km/h (rot, rechts).

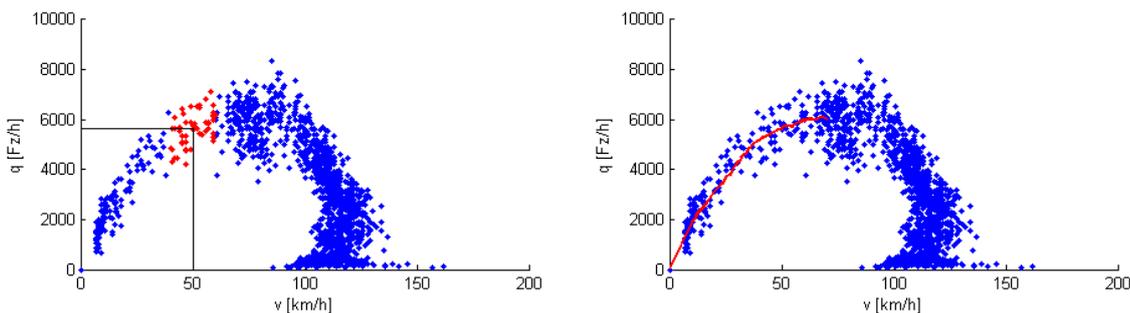


Abb. 4.1 Schätzung der Verkehrsstärke aus der Geschwindigkeit an einem lokalen Messquerschnitt mit Instanzbasiertem Lernen.

4.1.2 Anwendungsvoraussetzungen Instanzbasierter Lernverfahren

Zur erfolgreichen Anwendung eines Instanzbasierten Lernverfahrens müssen mehrere Voraussetzungen erfüllt sein. Grundvoraussetzung ist, dass ein Zusammenhang zwischen den unabhängigen Variablen X und der zur prognostizierenden abhängigen Variable y besteht.

Eine weitere Voraussetzung ist, dass die aktuellen Daten und die historischen Instanzen der gleichen Verteilung entstammen. Ein einfaches Beispiel kann anhand der Schätzung der lokalen Belegung aus der Tageszeit gezeigt werden. Wenn als zugrundeliegende Verteilung die Daten aller Tage genutzt werden, ist die Schätzung nicht optimal. Wenn ausschließlich die Daten einer falschen Tagesklasse genutzt werden (z.B. Sonntag statt Dienstag), ist die Schätzung falsch. Die beste Schätzung erhält man, wenn die Verteilung möglichst spezifisch ist (Abb. 4.2). Die Verteilung kann sich auch durch veränderte Randbedingungen ändern. Die typische Reisezeit beispielsweise kann sich durch eine Baumaßnahme von einem Tag auf den anderen ändern, so dass bestehende Instanzen zur Prognose auf einmal unbrauchbar werden.

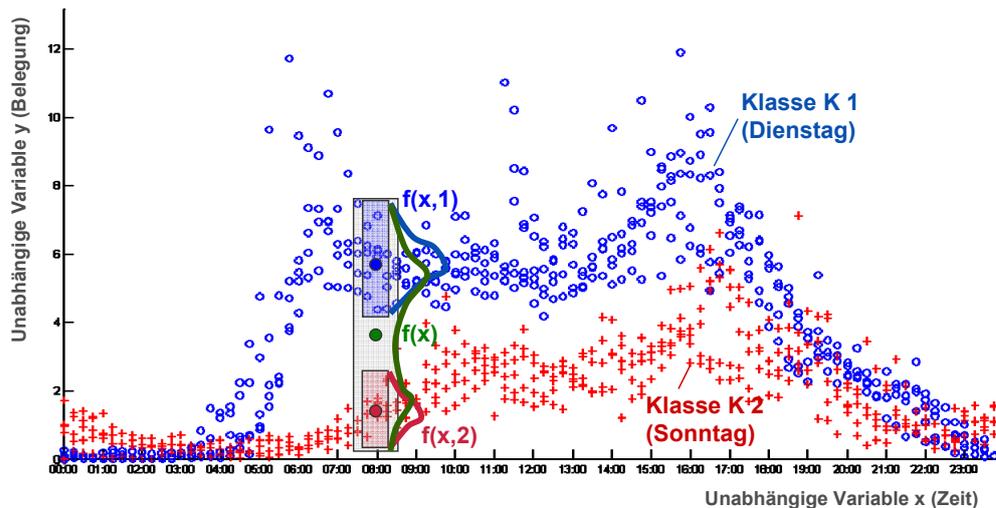


Abb. 4.2 Schätzung der Belegung aus der Tageszeit bei Vorliegen verschiedener Tagesklassen (Datenquelle: Graz, März/April 2007)

Der Merkmalsraum sollte möglichst vollständig von den Instanzen belegt sein, so dass die nächsten Nachbarn möglichst gleichmäßig um den aktuellen Merkmalsvektor verteilt liegen. Ansonsten kann es an den Rändern zu ungünstigen Schätzungen kommen. Abb. 4.3 zeigt ein Beispiel für einen eindimensionalen Merkmalsvektor. Für das links im Bild dargestellte x_i kann eine vermutlich gute Schätzung abgegeben werden. Das rechts im Bild dargestellte x_i liegt bezogen auf die Referenzbeobachtungen am Rand, so dass die Schätzung möglicherweise nicht optimal ist. Um eine möglichst vollständige Datenbasis zu haben, sollte eine möglichst große Basis an historischen Beobachtungen vorliegen.

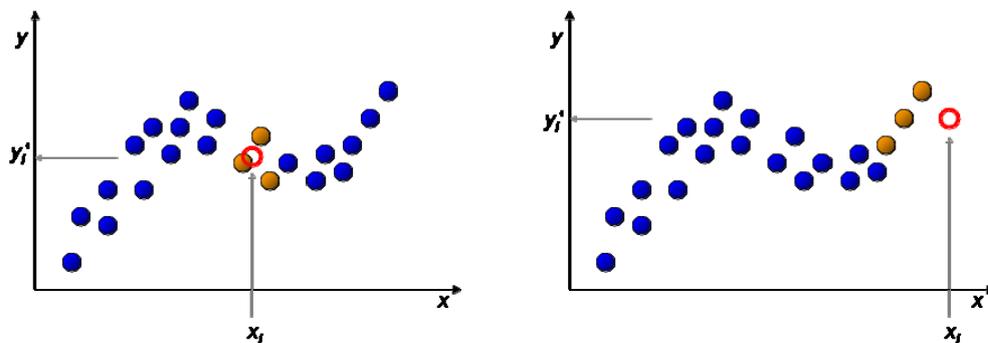


Abb. 4.3 Schätzung von $y' = f(x)$ mittels 3 Nächsten Nachbarn (3NN)

4.2 Auswahl der Merkmale und Definition der Distanzmetrik

4.2.1 Datenstruktur und Auswahl der Merkmale

Grundlage für das in diesem Kapitel beschriebene Verfahren ist eine strukturierte Datenbasis. Dabei wird unterschieden in kontinuierlich erfasste lokale Belegungsgrade, die als

unabhängige Prädiktorvariable genutzt werden, sowie in zufällig beobachtete Reisezeiten, die als zu schätzende abhängige Variable betrachtet werden. Bei der Prognose der Belegungsgrade werden diese selbst zu abhängigen Variablen. Die Reisezeiten der Routen r und die Belegungsgrade der Detektoren d werden in ein einheitliches Zeitraster von 15 Minuten gebracht. Jedes Zeitintervall besitzt darüber hinaus "Kalenderinformationen", also Datum, Zeit und Tagesklasse. Die Kalendervariablen werden bei der Prognose ebenfalls als Merkmal genutzt.

Wird zu einem Zeitpunkt t eine Prognose für Route r (gilt analog für einen Detektor d) für den Zeitpunkt $t + \Delta t$ gerechnet, dann werden zunächst diejenigen Detektoren d_r als mustergehend ausgewählt, die für die Route r relevant sind. Auswahlkriterium ist dabei der Abstand. Der Merkmalsvektor bzw. die Merkmalsmatrix $X_r(t+\Delta t)_{\text{aktuell}}$ setzt sich dann aus den Belegungsgraden der für r relevanten Detektoren für einen relevanten Zeitbereich $[t-dt_{\text{max}}, \dots, t]$ zusammen:

$$X_r(t + \Delta t) = \begin{bmatrix} x(t - dt_{\text{max}})_{r,\text{relevant } 1} & \dots & \dots & x(t - dt_{\text{max}})_{r,\text{relevant } n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x(t)_{r,\text{relevant } 1} & x(t)_{r,\text{relevant } 2} & \dots & x(t)_{r,\text{relevant } n} \end{bmatrix}$$

$X_i(t+\Delta t)_{\text{aktuell}}$ wird mit denjenigen historischen $X_i(t+\Delta t)_{\text{historisch}}$ verglichen, für die eine Beobachtung $y_i(t+\Delta t)_{\text{historisch}}$ vorliegt (siehe Abb. 4.4).

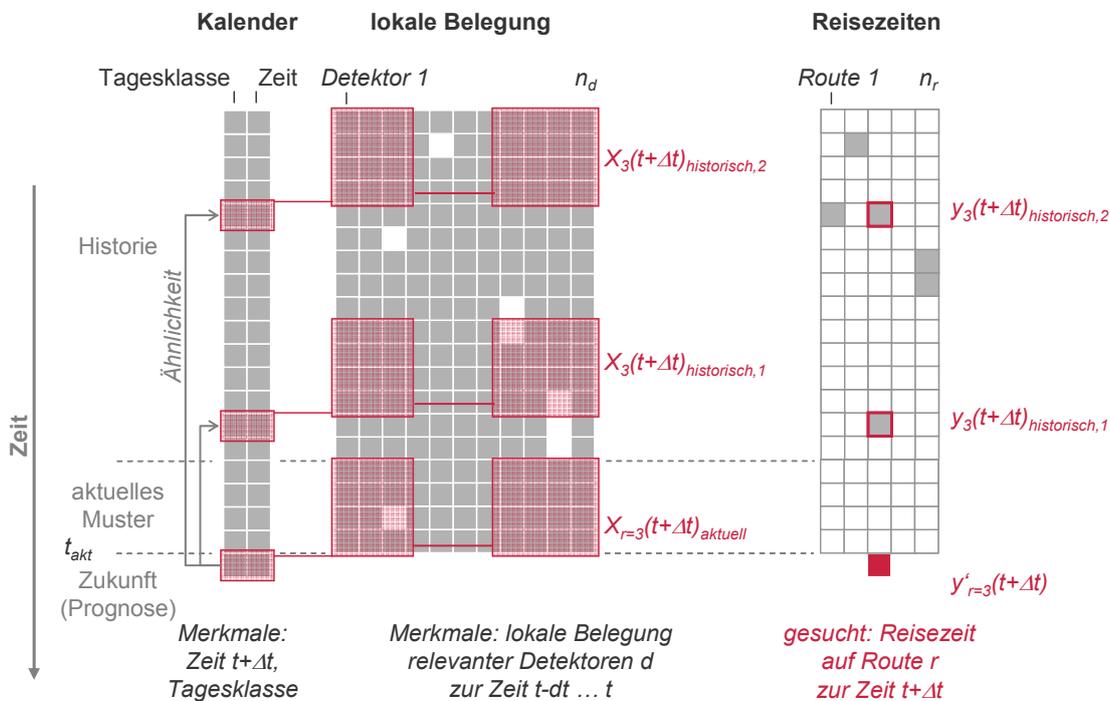


Abb. 4.4 Schematische Darstellung des Mustererkennungsprozesses

Bei dem Vergleich fließen neben den X_i der Verkehrsmerkmale auch Kalendermerkmale X_k mit ein. So werden historische Verkehrsmuster mit gleicher Ähnlichkeit bevorzugt, wenn sie ähnliche Kalendermerkmale besitzen. Fehlende $x \in X_i(t+\Delta t)$ werden über einen Strafterm berücksichtigt. Sämtliche Merkmale werden entsprechend der angenommenen Relevanz gewichtet.

Die zu den k ähnlichsten $X_i(t+\Delta t)_{\text{historisch}}$ gehörenden $y_i(t+\Delta t)_{\text{historisch}}$ werden zur Berechnung der Zielgröße $y'_i(t+\Delta t)$ genutzt. Die folgenden Abschnitte behandeln die dabei entstehenden Fragen zur Gewichtung der Merkmale entsprechend der räumlichen und zeitlichen Lage und zur expliziten Berücksichtigung unterschiedlicher Prognosehorizonte.

Im Folgenden wird eine Instanz (bestehend aus einem Muster und einer zugehörigen abhängigen Variable) mit dem Index i (mit $i = 1 \dots \text{Anzahl der Instanzen}$) und die Merkmale innerhalb eines Musters mit dem Index j (mit $j = 1 \dots \text{Anzahl der Merkmale}$) indiziert.

4.2.2 Berechnung der Distanz D zweier Verkehrssituationen

Die Distanz ("Unähnlichkeit") D zweier Verkehrssituationen (nicht Verkehrsmuster) berechnet sich allgemein zu:

$$D = G_V \cdot D_V + D_K + S_F$$

Dabei ist D_V die Distanz zwischen den eigentlichen Verkehrsmustern (bestehend aus lokalen Verkehrskenngrößen, vorzugsweise Belegungsgraden). D_K ist die Distanz aufgrund unterschiedlicher Kalendermerkmale und S_F ein Strafterm für fehlende Messwerte. G_V ist ein Parameter zur Gewichtung der Ähnlichkeit zweier Verkehrsmuster im Verhältnis zu D_K und S_F (vgl. Kapitel 5).

4.2.3 Distanz zweier Verkehrsmuster D_V

Abstandsmetrik

Die Ähnlichkeit zweier Muster oder Merkmalsvektoren kann auf unterschiedliche Weise definiert werden. Neben Korrelationsmaßen, die die Form des Vektors in den Vordergrund stellen, werden im Allgemeinen sogenannte Abstandsmetriken eingesetzt. Als Metrik bzw. Norm wird ein Abstandsmaß zwischen zwei Vektoren bezeichnet, das bestimmte Voraussetzungen erfüllt (siehe beispielsweise JUNG [2008]). Die so genannten *Minkowski Metriken* oder L_k *Normen* fassen die folgende Distanzfunktionsfamilie zusammen:

$$L_k\text{-Norm} \quad L_k = \left(\sum_{j=1}^m |x_j - x'_j|^k \right)^{\frac{1}{k}}$$

Bekannte Vertreter der L_k Normen sind diejenigen für $k = 1$ ("Manhattan Norm"), $k = 2$ ("Euklidische Distanz") und $k = \infty$ ("Maximum Norm" oder "Chebychev Distanz").

Manhattan-Norm
$$L_1 = \sum_{j=1}^m |x_j - x'_j|$$

Euklidische Distanz
$$L_2 = \sqrt{\sum_{j=1}^m |x_j - x'_j|^2}$$

Maximum-Norm
$$L_\infty = \left(\sum_{j=1}^m |x_j - x'_j|^\infty \right)^{\frac{1}{\infty}} \rightarrow \max_j \{|x_j - x'_j|\}$$

Mit einem großen Exponenten k (im Folgenden zur Vermeidung von Verwechslungen als k_e bezeichnet) wird dabei eine größere Gewichtung größerer Abstände erreicht, d.h. die Minimierung des maximalen Abstandes über alle Merkmale steht im Vordergrund. Die L_k Normen können um eine Gewichtung g_j der einzelnen Merkmale j erweitert und bezogen auf die Gesamtzahl der Merkmale m normiert werden:

gewichtete
normierte L_k -Norm
$$L_k = \frac{\left(\sum_{j=1}^m |(x_j - x'_j) \cdot g_j|^{k_e} \right)^{\frac{1}{k_e}}}{m}$$

Die Faktoren g_j erlauben, jedes Merkmal entsprechend seiner räumlichen und zeitlichen Lage zu gewichten. Die Normierung wird zur Berücksichtigung unterschiedlich langer Merkmalsvektoren aufgrund möglicherweise fehlender Messwerte benötigt.

Räumlich-zeitliche Gewichtung der Messwerte

Grundlegende Annahme ist, dass eine Prognose an einem Ort r zum Zeitpunkt t nicht nur auf der aktuellen lokalen Situation in Form des Messwertes $x(r,t)$ basieren kann und sollte, sondern dass durch die Berücksichtigung erweiterter räumlicher und zeitlicher Information die Situation und deren zukünftige Entwicklung besser charakterisiert werden kann.

Die Eignung einer Verkehrskenngröße am Ort dr als Prädiktor für eine Verkehrskenngröße am Ort r hängt innerorts im Wesentlichen von der Verteilung der Nachfrage im Netz ab. Je größer der Anteil der dr passierenden Fahrzeuge auch r passieren, desto höher die Prädiktorqualität von dr für r . Konsequenterweise wäre das exakte Wissen der von dr ausgehenden Routenspinnen, z.B. aus einer plausiblen dynamischen Umlegung, ein guter Indikator zur Schätzung der Prädiktorqualität von dr für r . Häufig – so auch im vorliegenden Fall – lie-

gen keine Routenspinnen vor, so dass die Prädiktorqualität auf andere Weise ermittelt bzw. abgeschätzt werden muss. Bei vorliegenden Zeitreihen $x(r)$ und $x(dr)$ könnte eine Korrelationsanalyse Aufschluss über den Einfluss von dr auf r geben. Hier besteht jedoch die Problematik, dass Korrelationen aufgrund täglich wiederkehrender Nachfragemuster sehr ausgeprägt sind. Beispielsweise weisen Messstellen auf Einfallstraßen aus Norden und Süden hohe Korrelationen auf, jedoch wird die eine über die Bestätigung täglich wiederkehrender Muster hinaus kaum als Prädiktor für die andere dienen können. Als pragmatischer Ersatz für eine Routenspinne wird hier der Abstand von dr nach r als Indikator für die Prädiktorqualität von dr nach r verwendet. Der Abstand $ds_{r,dr}$ ist dabei der kürzeste Weg im Straßennetz von dr nach r .

Der Abstand $ds_{r,dr}$ wird null, wenn die Orte dr und r identisch sind (Messstelle bei der Prognose lokaler Verkehrskenngrößen) oder dr auf r liegt (Messstelle liegt direkt auf der Route, für die eine Verkehrskenngröße prognostiziert werden soll).

Innerorts wird davon ausgegangen, dass der Einfluss von dr auf r vom Abstand von dr nach r in Flussrichtung abhängt. Diese Annahme basiert auf der Tatsache, dass sich die Information in Form der Nachfrage innerorts primär in Flussrichtung ausbreitet und dass sich relevante Ereignisse wie Staus auch primär durch eine zu hohe Nachfrage an den Knotenpunkten einstellen. Zur Berücksichtigung von Rückstaus, die von dr nach r wandern können, müsste naturgemäß der Abstand von dr nach r entgegen der Flussrichtung verwendet werden. Dies wäre insbesondere auf Schnellstraßen relevant, wo Stauwellen häufig über mehrere Kilometer stromaufwärts wandern.

Neben räumlich entfernten Messwerten wird auch zeitlich entfernten, also in der Vergangenheit liegenden, Messwerten, eine Prädiktoreigenschaft zugeschrieben. In einem dynamischen System wie dem Verkehrsablauf in einem Straßennetz ist nicht nur die aktuelle Ausprägung einer Messgröße von Belang, sondern auch wie diese entstanden ist. Dieselbe Belegung kann beispielsweise zu einer zu- oder abnehmenden Zeitreihe gehören, was für die Prognose sicherlich relevant ist. Auch kann der Einfluss von trotz Aggregation und Glättung auftretenden Messwerteausreißern reduziert werden. Es wird angenommen, dass die Belegungsgrade von Messstellen dr im Netz für Zeitintervalle $t-dt$ für $dt = [0, \dots, dt_{max}]$ dienen können.

Die Relevanz eines Merkmals $x(ds_{r,dr}, dt)$ wird dabei durch das Gewicht $g(ds_{r,dr}, dt)$ bei der Berechnung der Musterähnlichkeiten berücksichtigt (siehe Abb. 4.5).

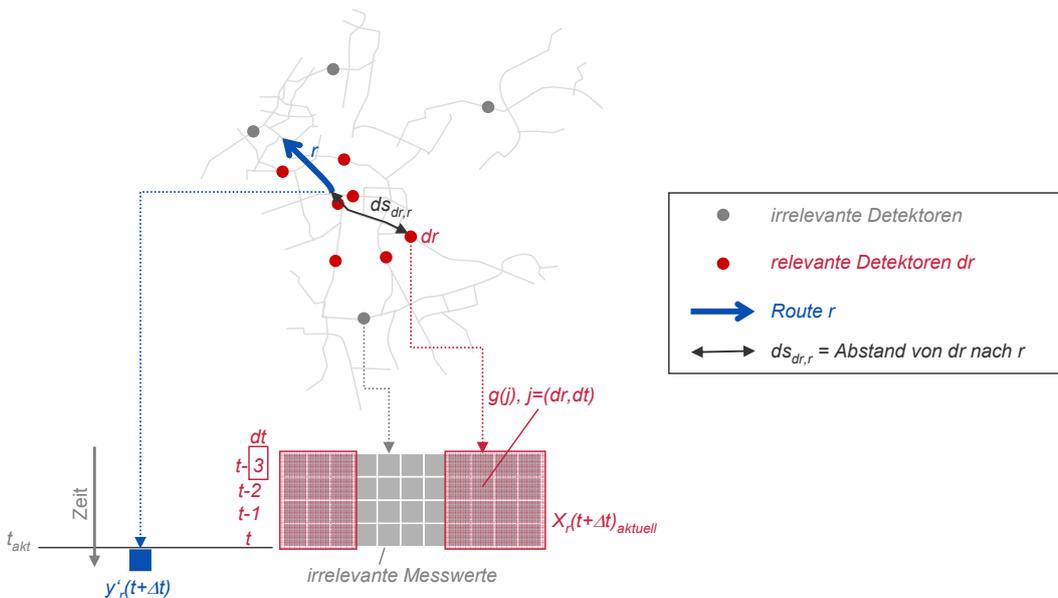


Abb. 4.5 Darstellung der Merkmalsgewichtung als Funktion von Ort und Zeit

Der Gewichtungsfaktor $g(ds_{r,dr}, dt)$ wird als lineare Funktion $g = f(ds_{r,dr}, dt)$ in Abhängigkeit von den Parametern, ds_{max} und dt_{max} formuliert zu:

$$g(ds_{r,dr}, dt) = \max \left\{ 0; 1 - \frac{ds_{r,dr}}{ds_{max}} - \frac{dt}{dt_{max}} \right\}$$

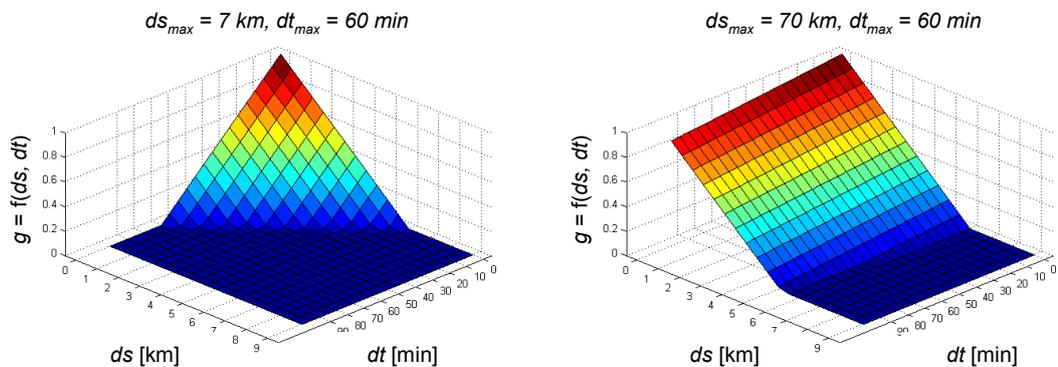


Abb. 4.6 Gewichtungsfunktionen g für verschiedene Parameter ds_{max} und dt_{max} (dt ist der Interpretierbarkeit wegen in Minuten dargestellt)

Mit den Parametern ds_{max} und dt_{max} kann eingestellt werden, wie schnell das Gewicht mit steigenden ds (Abstand der Messstelle) und dt (Zeitdifferenz zwischen dem aktuellen und dem zum Messwert gehörenden Zeitintervall) fällt.

Die lineare Form der Gewichtungsfunktion kann prinzipiell durch andere Formen (z.B. Exponentialfunktion) ersetzt werden. Die Distanz $D_{V,i}$ zwischen dem aktuellen Verkehrsmuster X_i^{akt}

und dem historischen Verkehrsmuster X_i berechnet sich als gewichteter mittlerer Abstand bezogen auf den Mittelwert \bar{x}_i^{akt} über dt und j des aktuellen Musters:

$$D_{V,i} = \frac{\left(\sum_{dt=0}^{dt_{max}} \sum_{j=1}^m \left(g(ds_{r,dr}, dt) \cdot |x_{i,j}^{akt}(dt) - x_{i,j}^{hist}(dt)| \right)^{k_e} \right)^{\frac{1}{k_e}}}{\left(\sum_{dt=0}^{dt_{max}} \sum_{j=1}^m g(ds_{r,dr}, dt) \right) \cdot \bar{x}_i^{akt}} \quad \text{mit } 0 \leq g \leq 1$$

Die Annahme, dass die Gewichtung mit zeitlichem und räumlichem Abstand abnimmt, vernachlässigt die Fortpflanzungsgeschwindigkeit der Information im Straßennetz. In Flussrichtung ist dies die Reisegeschwindigkeit, entgegen der Flussrichtung wäre es die Stauausbreitungsgeschwindigkeit. Verkehrsdynamisch intuitiv wäre es, das Gewicht nicht bei $ds = 0$ und $dt = 0$ maximal werden zu lassen, sondern das Maximum entsprechend der (Informations-)Reisezeit zu setzen. So wäre für einen Detektor d beispielsweise zu erwarten, dass eine 5 km stromaufwärts beobachtete Nachfragespitze bei einer Reisegeschwindigkeit von 25 km/h nach 0.2 h = 12 Minuten (mehr oder weniger verzerrt) an d beobachtet werden kann und somit g bei 5 km und 12 Minuten maximal ist. Aufgrund der Aggregation auf 15-Minutenintervalle und da die berücksichtigten Detektoren meist in einem Abstand von weniger als 5 km liegen, kann die Information normalerweise innerhalb eines Zeitintervalls zum Ort der Prognose reisen. Daher wird die Gewichtungsfunktion hier vereinfacht mit $g_{max} = 1$ für $ds = 0$ und $dt = 0$ für jeden Detektor angenommen.

4.2.4 Kalenderbasierte Distanz D_K

Kalendermerkmale (Tagesklasse und Tageszeit) stellen aufgrund regelmäßiger Nachfragestrukturen wichtige Merkmale zur Prognose verkehrlicher Kenngrößen dar. Daher werden die Ähnlichkeit der aktuellen und der historischen Tagesklasse, sowie die Übereinstimmung der Tageszeiten mit als Merkmal zur Bewertung der Ähnlichkeit zweier Situationen angenommen.

Die Tagesklassen werden anhand der Wochentage und weiterer verkehrsrelevanter Merkmale wie "Feiertag" und "Brückentag" festgelegt. In einem Ereigniskalender können einzelne Tage zusätzlich entsprechend verkehrswirksamer Ereignisse (z.B. Münchner Bladenight) definiert werden. Die Ähnlichkeit zweier Tagesklassen ist im Wesentlichen von der Nachfragestruktur abhängig, da mögliche variable Angebotseigenschaften – wie z.B. LSA-Programme – an die Nachfragestruktur angepasst oder sogar verkehrsadaptiv und somit implizit enthalten sind.

Zur Beschreibung der Unähnlichkeit des Merkmals (*Tagesklasse, Tageszeit*) des aktuellen Musters zu einer historischen Instanz i wird das Distanzmaß $D_{K,i}$ definiert. Dabei soll $D_{K,i}$ klein

sein, wenn sich der Erwartungswert der zu prognostizierenden Größe für das Merkmal (*Tagesklasse, Tageszeit*) für die Muster wenig unterscheidet. $D_{K,i}$ wird berechnet zu:

$$D_{K,i}((Tageskl_i, Tagesz_i), (Tageskl_{akt}, Tagesz_{akt})) = \frac{E_y(T_i, Z_i) - E_y(T_{akt}, Z_{akt})}{\bar{y}}$$

E_y ist dabei der Erwartungswert der abhängigen Variablen je Tagesklasse und Zeitintervall. Durch Division durch den Mittelwert (Erwartungswert) der abhängigen Variablen über die gesamte Stichprobe \bar{y} werden die Unterschiede in den Erwartungswerten relativiert. Dabei sind in der "Grundversorgung" sieben Tagesklassen (jeder Wochentag, Feiertage werden in die Sonntagsklasse sortiert) und 96 Zeitintervalle (15 Minuten) je Tag vorgesehen.

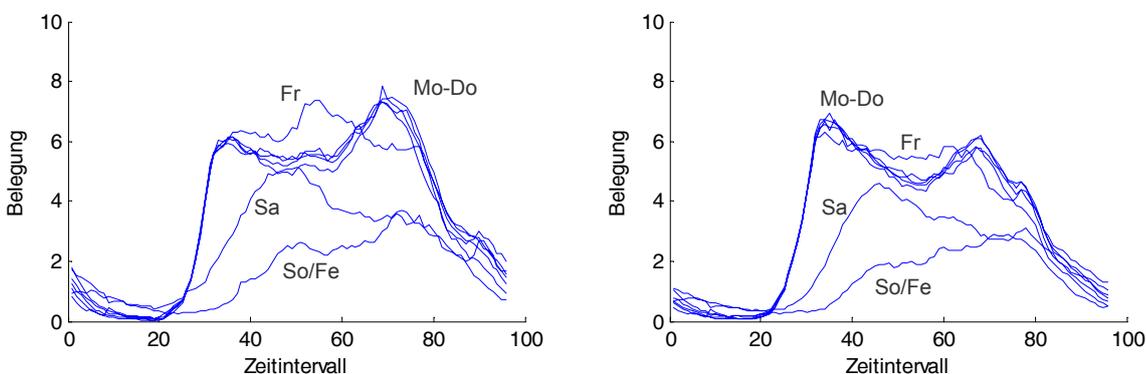


Abb. 4.7 Darstellung der Repräsentanten für verschiedene Tagesklassen für zwei Detektoren in Graz (Zeitraum: März 2007 – Mai 2008)

Eine Analyse zeigt (wie vermutet), dass die Ähnlichkeit zwischen den Tagesklassen teilweise sehr groß ist, die entsprechenden $D_{K,i}$ für dieselben Zeitintervalle also sehr klein sind. Abb. 4.7 zeigt Beispiele für die Erwartungswerte des Belegungsgrades über die 96 Zeitintervalle für verschiedene Tagesklassen für zwei Detektoren. Es ist deutlich zu sehen, dass Montag bis Donnerstag sehr ähnlich verlaufen. Dementsprechend werden bei der Prognose des Erwartungswertes an einem Dienstag für 09:00 Uhr die historischen 09:00 Uhr-Instanzen der Montage, Dienstag, Mittwoch und Donnerstag quasi gleichberechtigt behandelt.

Durch die Integration der Tagesklassen und der Tageszeit als Distanzfunktion wird eine Berücksichtigung dieser insbesondere für größere Prognosehorizonte relevanten Merkmale möglich. Gleichzeitig ist durch eine entsprechende Gewichtung auch das Tagesklassen- und Tageszeit-übergreifende Erkennen signifikanter Verkehrssituationen möglich, da der Mustervergleich nicht auf identische Tagesklassen beschränkt ist. Liegt zum Beispiel an einem Montag um 09:00 Uhr eine besondere Verkehrsstörung – charakterisiert durch ein signifikantes Muster in den betreffenden Verkehrsdaten – vor, so kann die Betrachtung ähnlicher historischer Verkehrsdatenmuster eines Freitags um 12:00 Uhr zur Prognose für kurze Horizonte oder zur Schätzung der aktuellen Reisezeit sinnvoller sein, als die Betrachtung des ähnlichsten Musters derselben Tagesklasse aus demselben Zeitintervall. Andererseits

sollte bei zwei bezüglich des Verkehrsmusters ähnlichen historischen Instanzen derjenigen der Vorzug gegeben werden, die bezüglich der Kalendermerkmale besser zur aktuellen Situation passt. Die so für Graz ermittelten Distanzen D_K sind für zwei Beispiele in Abb. 4.8 dargestellt.

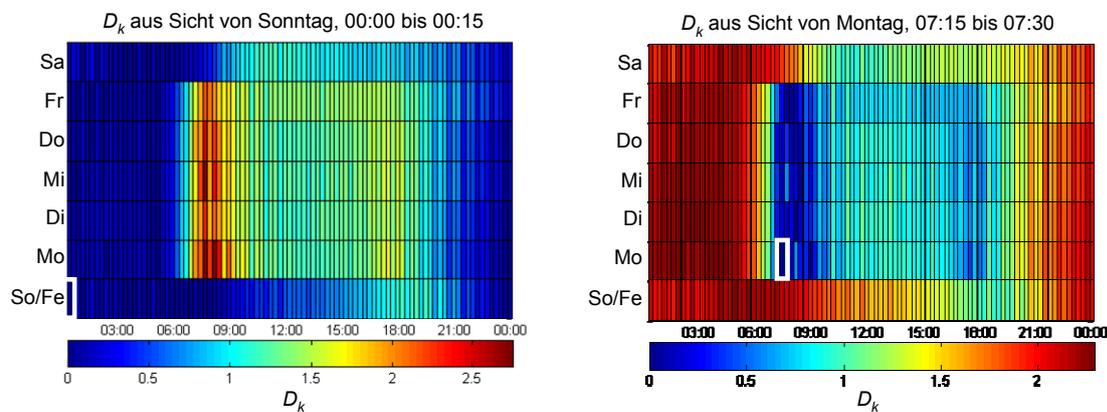


Abb. 4.8 Darstellung der Distanzen D_K für sonntags 00:00-00:15 Uhr (links) und montags 07:00-07:30 Uhr (rechts), Datenbasis: Belegungsgrade "EXERZIERPLATZ"

Links sind die Distanzen D_K aus Sicht von Sonntag und dem Zeitintervall 00:00 Uhr bis 00:15 Uhr dargestellt. D_K ist für die Morgenspitzen von Montag bis Freitag am höchsten. Umgekehrt verhält es sich entsprechend aus Sicht von Montag im Zeitintervall 07:00 Uhr bis 07:15 Uhr (rechts).

4.2.5 Strafterm S_F zur Berücksichtigung fehlender Messwerte

Wie in Abschnitt 2.1.3 beschrieben, kann es aufgrund fehlender oder unplausibler Messwerte zu Löchern in der eigentlich kontinuierlichen lokalen Datenerfassung kommen. Dadurch können sowohl das aktuelle als auch die historischen Verkehrsmuster unvollständig sein. Muster sollten jedoch auch bei fehlenden Messwerten bezüglich ihrer Ähnlichkeit bewertet werden können.

Die in Abschnitt 4.2.3 beschriebene Funktion zur Berechnung der Ähnlichkeit D_V zweier Verkehrsmuster ist normiert auf die Anzahl n der verwendbaren Muster. Ergebnis ist also eine durchschnittliche Distanz der einzelnen Merkmale. Eine gleiche Ähnlichkeit der Verkehrsmuster sollte jedoch höher bewertet werden, wenn das Muster vollständig ist. Ansonsten besteht die Gefahr, dass stark unvollständige Muster mit einzelnen zufällig ähnlichen Messwerten als nächste Nachbarn identifiziert werden.

Bei der Distanzberechnung wird daher der Strafterm S_F für fehlende Messwerte eingeführt. S_F berechnet sich zu:

$$S_{F,i} = \frac{n_{gesamt,i} - n_{gültig,i}}{n_{gesamt,i}} \cdot \left(1 - \frac{t_{prog}}{t_{prog,max}} \right)$$

mit $0 \leq t_{prog} \leq t_{prog,max}$ und $0 \leq n_{gültig,i} \leq n_{gesamt,i}$

Dabei ist $n_{gesamt,i}$ die Anzahl der möglichen Verkehrsmerkmale des Musters X_i . $n_{gültig,i}$ ist die Anzahl der Merkmale, die sowohl im aktuellen als auch im historischen Muster X_i vorhanden sind (siehe Abb. 4.9).

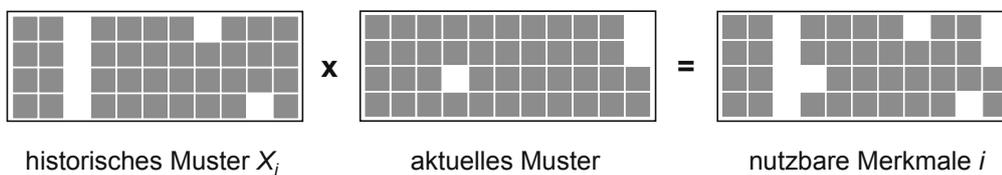


Abb. 4.9 Ermittlung der Anzahl der gültigen Merkmale des Verkehrsmusters

$S_{F,i}$ nimmt dabei, proportional mit steigendem Prognosehorizont ab, ebenso wie die Gewichtung der Ähnlichkeit der Verkehrsmuster

4.2.6 Berücksichtigung weiterer Merkmale

Das Verfahren der Ähnlichkeitsberechnung ist derart konzipiert, dass potentielle weitere relevante Merkmale M einfach über einen zusätzlichen Distanzterm D_M und einen Gewichtungsfaktor G_M integriert werden können. Relevante Merkmale könnten beispielsweise Wetter- oder Veranstaltungsinformationen sein.

4.3 Anzahl Nachbarn k und lokale Approximation der Daten

Durch die im vorherigen Abschnitt definierte Abstandsmetrik wird die Distanz des aktuellen Musters zu allen Instanzen berechnet. Sortiert man die Instanzen aufsteigend nach der Distanz, so handelt es sich bei den k ersten Listeneinträgen um die k nächsten Nachbarn. Wesentliche Verfahrensparameter sind die Anzahl der zu verwendeten Nachbarn k und das Verfahren zur lokalen Funktionsapproximation aus den k nächsten Nachbarn.

4.3.1 Anzahl der zu berücksichtigenden Nachbarn k

Die Anzahl der zu berücksichtigenden Nachbarn k bestimmt Glättung und Spezifität. Je größer k , desto stärker geglättet und weniger spezifisch ist die Schätzung für y . Wird im Extremfall k gleich der Anzahl der zur Verfügung stehenden Beobachtungen gewählt, so ist die

Schätzung für y gleich dem Mittelwert über alle jemals gemachten Beobachtungen gewichtet mit der Ähnlichkeit der jeweiligen Instanz. Wird k sehr klein gewählt (z.B. $k = 1$), so ergibt sich die Schätzung auf Basis genau einer Beobachtung und ist somit zufälligen Schwankungen "ausgeliefert".

4.3.2 Lokale Approximationsfunktion / Gewichtungsfunktion

Die lokale Approximation der Daten beschreibt den Prozess der Ermittlung der zu schätzenden Größe.

Die k nächsten Nachbarn $[X_i, y_i]_{i=1 \dots k}$ mit zugehöriger Distanz D_i zum aktuellen Verkehrsmuster werden zur Prognose herangezogen. Die lokale Approximationsfunktion bestimmt dabei, wie das zu X_{aktuell} gehörige y'_{aktuell} aus $[X_i, y_i]$ und D_i geschätzt wird.

Das vorliegende Problem kann wiederum als Regressionsproblem aufgefasst werden. Nur werden jetzt nur die k relevantesten Instanzen verwendet und nicht die gesamte Datenbasis. Es wird angenommen, dass die k nächsten Nachbarn ähnlich sind, so dass typischerweise relativ konservative und einfache Funktionen zur lokalen Approximation verwendet werden. Allgemein wird der Einfluss der Nachbarn als Funktion ihrer Distanz zum aktuellen Muster mit einer sogenannten Kernfunktion K berechnet.

Dabei ist $K(X_{\text{aktuell}}, X_{\text{Instanz}})$ abhängig von der Distanz D der beiden Muster X_{aktuell} und X_{Instanz} . Für K kommen verschiedene Funktionsformen in Frage. Gebräuchlich sind Gauß-Kerne, Norm-Kerne und Kerne quadratischer Form. Wird $K=\text{konstant}$ gewählt, so haben alle berücksichtigten Instanzen unabhängig von ihrer Ähnlichkeit mit dem aktuellen Muster das gleiche Gewicht. In Abb. 4.10 sind verschiedene Funktionsformen für die Funktion K dargestellt.

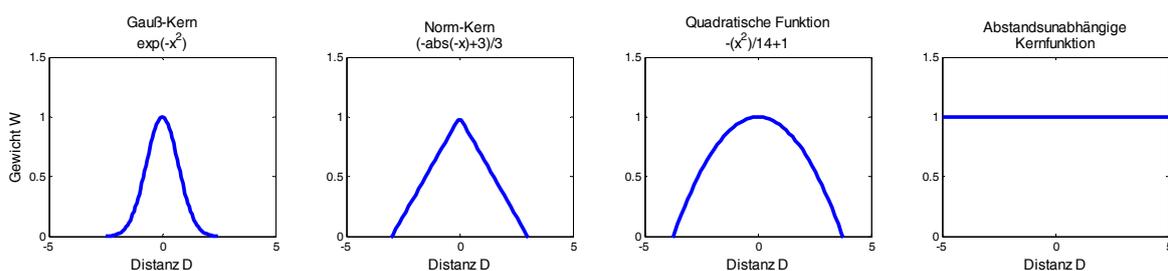


Abb. 4.10 Darstellung verschiedener Kernfunktionsformen K

In dieser Arbeit wird der Gauß-Kern verwendet, da er einige günstige Eigenschaften aufweist: Das Gewicht, also der Einfluss, einer Instanz konvergiert mit steigender Distanz exponentiell abnehmend gegen null. Somit werden Instanzen mit geringer Distanz deutlich bevorzugt. Gleichzeitig wird das Gewicht aber nie gleich null, so dass auch dann eine Lösung ermittelt werden kann, wenn alle Instanzen eine relativ große Distanz zum aktuellen

Muster aufweisen. Des Weiteren besitzt der Gauß-Kern mit der Bandbreite K_W nur einen einzustellenden Parameter. Das Gewicht w_i der Instanz i berechnet sich zu:

$$w_i = e^{-\frac{D_i/D_{i,\min}}{K_w}}$$

K_W ist dabei die sogenannte Kernweite (*engl.: Kernel Width*) und bestimmt den Einfluss der berücksichtigten Instanzen als Funktion ihrer Distanz zum aktuellen Muster. Wird K_W groß gewählt, dann haben auch unähnliche Instanzen einen relativ großen Einfluss auf die Prognose und die Prognose wird konservativer. Die Distanz D_i wird auf das minimale $D_{i, i=1\dots k}$ bezogen, so dass die Distanz der ähnlichsten Instanz auf den Wert 1 normiert ist.

Der Funktionswert y'_{aktuell} für X_{aktuell} wird nun aus den gewichteten Instanzen berechnet. Meist wird dazu der gewichtete Mittelwert (*Verfahren der gewichteten nächsten Nachbarn oder Kernregression, engl.: Kernel Regression*), eine gewichtete lineare Regression (*lokale gewichtete Regression, engl.: Locally Weighted Regression*) oder ein anderer "konservativer" Schätzer verwendet, da die Nachbarn bei einer sinnvoll gewählten Distanzfunktion D die Funktionswerte y' an sich schon gut annähern sollten.

Während die Form der lokalen Approximation bei einer gleichmäßigen Verteilung der Instanzen um den aktuellen Merkmalsvektor einen relativ geringen Einfluss hat, so kann der Unterschied in den Randbereichen beträchtlich sein. ROBINSON [2005] bemerkt dazu: "*One criticism of the use of the mean, median, and weighted mean, is that these estimators are susceptible to boundary bias. The k nearest points may tend to lie on one side of the input feature vector, thus giving a biased result. An alternative local estimation method is to use the k nearest neighbors to determine a linear regression model suitable for the local area.*"

Die Tatsache, dass die Schätzung mit Mittelwert und Median an den Rändern möglicherweise verzerrt ist, lässt sich gut an einem einfachen Beispiel mit einer unabhängigen Variablen zeigen. Abb. 4.11 zeigt beispielhaft den Fall, dass der Eingangsvektor nicht von Merkmalsvektoren der historischen Musterdatenbasis umgeben ist, sondern außerhalb liegt. Andererseits ist es fraglich, ob eine lineare Regression oder irgendeine andere Funktionsform zu einer verbesserten Schätzung führen. Im untenstehenden Beispiel würde eine lokal gewichtete Regression dazu führen, dass bei einer Belegung von etwa 67% eine Reisegeschwindigkeit von 0 km/h geschätzt würde. Mittels Kernregression wäre die Schätzung 12 km/h. Das Problem ist, dass der funktionale Zusammenhang zwischen der Belegung und der Reisezeit nicht bekannt ist (was einer der Hauptgründe dafür ist, dass ein Instanzbasiertes Lernverfahren überhaupt eingesetzt wird).

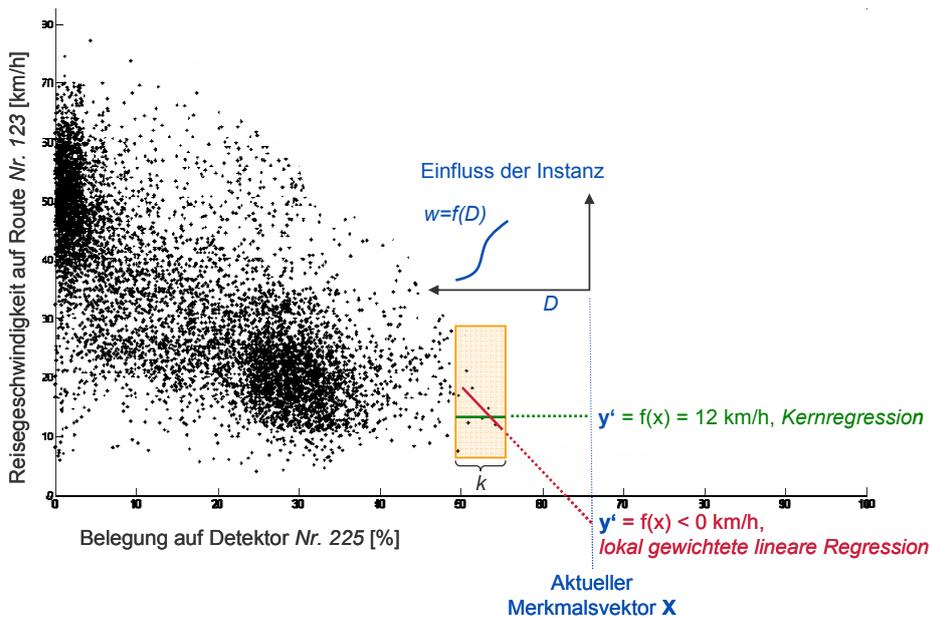


Abb. 4.11 Reisegeschwindigkeit auf einer Route als Funktion des Belegungsgrades eines Detektors

Ein Grund, der für die Nutzung sehr konservativer Schätzer (Kernregression, Median der nächsten Nachbarn) spricht ist, dass die Schätzung innerhalb eines schon beobachteten Wertebereichs liegt. Z.B. kann die Reisezeit niemals höher geschätzt werden als der maximal jemals beobachtete Wert. Dies erscheint insbesondere bei hochdimensionalen Merkmalsvektoren (beispielsweise 10 Detektoren \times 4 Zeitintervalle) sinnvoll, da die Auswirkung der Lage einzelner Merkmale nicht mehr einfach abgeschätzt und nicht visualisiert werden kann. Darüber hinaus sind konservative Schätzer robuster gegenüber Ausreißern, so dass im Folgenden die Kernbasierte Regression als lokale Approximationsmethode verwendet wird.

4.4 Berücksichtigung aktueller Messwerte der Zielgröße

Das Verfahren geht im Allgemeinen davon aus, dass die zu prognostizierende Größe zufällig beobachtet wird und somit für den aktuellen Zeitpunkt t nicht vorliegt. Liegt jedoch eine geschlossene Zeitreihe bis zum aktuellen Zeitpunkt t vor (dies ist in der Regel bei der lokalen Verkehrsdatenerfassung der Fall), so ist für kurze Prognosehorizonte Δt eine verbesserte Prognose durch explizite Berücksichtigung der aktuellen Messung zu erwarten. Dieses Vorgehen ist in der Praxis und in der Forschung bei der Prognose lokaler kontinuierlich erfasster Verkehrskenngrößen gängig [WILD 1996, VORTISCH 2005]. Prinzipiell können dabei zwei Ansätze angewendet werden: Das Angleichen des aktuellen Messwertes an den prognostizierten Wert und die Prognose auf Basis des aktuellen Messwertes und des erwarteten Inkrements. Beide verhalten sich bezüglich der Wirkung ähnlich, benötigen jedoch unterschiedliche Eingangsdaten.

4.4.1 Angleichen an den aktuellen Messwert

Mit dem vorgestellten Verfahren können Prognosen für verschiedene Horizonte als wahrscheinlichste Ergebnisse aus den historischen Beobachtungen ermittelt werden. Es kann vorkommen, dass das Niveau der aktuellen Ganglinie über oder unter dem der Prognose für den aktuellen Zeitpunkt (Horizont = 0 Minuten) liegt. In diesem Fall ist zu erwarten, dass auch die Prognose für die nächsten Intervalle (15 Minuten, 30 Minuten) verzerrt ist. Diese Verzerrung kann mit Hilfe eines Angleichfaktors η eliminiert oder zumindest reduziert werden. Dabei wird die Prognose für den Horizont Δt aus der ursprünglichen Prognose $y'(t+\Delta t)$ und der Differenz zwischen dem aktuellen Messwert $y(t)$ und der Prognose $y'(t)$ berechnet und über η angepasst:

$$y'_{\text{angepasst}}(t + \Delta t) = y'(t + \Delta t) + \eta \cdot (y'(t) - y(t))$$

Dabei ist $\eta = \max\{0, 1 - \Delta t / \Delta t_{\max}\}$ ein mit dem Prognosehorizont abnehmender Faktor. Dadurch werden zeitnahe Werte stärker von der aktuellen Messung beeinflusst. Ab einem zu definierenden Horizont Δt_{\max} (hier: 1 Stunde) verschwindet der Einfluss der aktuellen Messung.

4.4.2 Inkrementprognose

Statt der zu prognostizierenden Zielgröße $y(t+\Delta t)$ können die historisch beobachteten Inkremente $\Delta y(t, t+\Delta t) = y(t+\Delta t) - y(t)$ als Zielgröße gespeichert werden. Die Prognose berechnet sich dann aus dem aktuellen Messwert $y(t)$ und dem erwarteten Inkrement $\Delta y(t, t+\Delta t)$ zu:

$$y'_{\text{angepasst}}(t + \Delta t) = y(t) + \Delta y'(t + \Delta t)$$

Die Inkrementprognose hat den Vorteil, dass das Niveau des aktuellen Werts automatisch berücksichtigt wird. Die Inkrementprognose eignet sich nur für Prognosehorizonte, in denen ein Einfluss des aktuellen Messwertes unterstellt werden kann (also etwa 15 bis maximal 30 Minuten). Der Nachteil gegenüber dem Angleich-Verfahren aus dem vorhergehenden Abschnitt ist, dass kurze Messwertspitzen ungedämpft in die Zukunft extrapoliert werden, was bei kurzen Horizonten richtig sein mag, bei größeren Horizonten aber zu Überschätzungen der Zielgröße führen dürfte.

Obgleich die Inkrementprognose intuitiv ansprechend erscheint, haben Tests gezeigt, dass das "Angleichen an den aktuellen Messwert" für die hier entwickelte Anwendung besser geeignet ist. Daher wird die Angleich-Methode zur Berücksichtigung der aktuellen Zeitreihe in der Prognose genutzt (siehe Abschnitt 5.4.1).

4.5 Prognosezuverlässigkeit

Die Information über die Zuverlässigkeit einer Prognose kann ebenso bedeutsam sein wie die Prognose des Erwartungswertes selbst.

Die Zuverlässigkeit kann dabei prinzipiell auf zwei Weisen qualifiziert bzw. quantifiziert werden. Die erste Möglichkeit ist die Angabe von Zuverlässigkeitsindikatoren. Die Indikatoren sollten dem Nutzer einen Rückschluss auf die Erwartungstreue der Prognose erlauben. Die zweite Möglichkeit ist die Konstruktion von Vertrauensintervallen, also die Angabe von Korridoren, innerhalb derer die Prognosegröße vermutlich mit einer bestimmten Wahrscheinlichkeit liegen wird. Das Instanzbasierte Lernen erlaubt aufgrund seiner Struktur eine a priori Abschätzung der Zuverlässigkeit der Prognose auf Basis der Musterähnlichkeiten und der Verteilung der zu den ähnlichsten Instanzen gehörenden abhängigen Variablen. Folgende Eigenschaften werden angenommen:

- Je näher das aktuelle Muster zu seinen nächsten Nachbarn liegt, desto zuverlässiger ist die Prognose.
- Je gleichmäßiger die ähnlichsten Nachbarn um das aktuelle Muster herum verteilt sind, desto zuverlässiger ist die Prognose.
- Je mehr Nachbarn mit einer hohen Ähnlichkeit gefunden werden, desto zuverlässiger ist die Prognose.
- Je geringer die Streuung innerhalb der zu den nächsten Nachbarn gehörenden abhängigen Variablen ist, desto zuverlässiger ist die Prognose.
- Zur Erklärung wird die Problematik anhand eines einfachen zweidimensionalen Beispiels dargestellt.

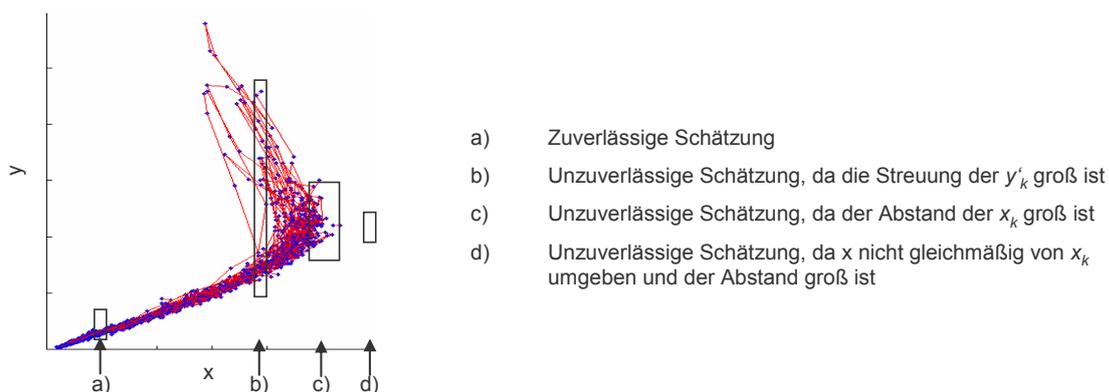


Abb. 4.12 Darstellung unterschiedlich zuverlässiger Schätzungen

Die Zuverlässigkeitsindikatoren werden daraus abgeleitet im Folgenden formuliert.

4.5.1 Zuverlässigkeitsindikatoren

Zur Zuverlässigkeitsabschätzung infragekommene verfahrensimmanente Größen sind (1) die Lage des Verkehrsmusters X_{aktuell} im Vergleich zu seinen nächsten Nachbarn in der historischen Datenbasis, (2) die tatsächliche Ähnlichkeit zwischen X_{aktuell} im Vergleich zu seinen nächsten Nachbarn in der historischen Datenbasis und (3) die Verteilung der zur Prognose genutzten k abhängigen $y_{\text{historisch}}$.

Die Prognosezuverlässigkeit wird typischerweise für relevante Prognosen (z.B. Reisezeit bei einem nicht wiederkehrenden Stau) sehr viel unzuverlässiger sein als für weniger relevante (z.B. Belegungsgrad nachts um 02:00 Uhr). Während das intuitiv angenommen wird, so lässt es sich auch durch die Tatsache erklären, dass im ersten Fall X_{aktuell} vermutlich eher am Rand liegt, die tatsächliche Ähnlichkeit zu den vorhandenen Instanzen klein ist (wenn noch keine derartige Situation beobachtet wurde) und dass die Verteilung der k ähnlichsten $y_{\text{historisch}}$ vermutlich eine große Streuung aufweist. Für den zweiten Fall ist jeweils das Gegenteil der Fall.

Lage von X_{aktuell}

Die Lage L_X eines Merkmalsvektors X_{aktuell} zu den k nächsten historischen Instanzen $X_i^{\text{historisch}}$ wird ermittelt zu:

$$L_X = \text{Median}\left(X_i^{\text{historisch}} - X_{\text{aktuell}}\right)_{i=\text{Instanz } 1 \dots k}$$

Typischerweise konzentrieren sich die Merkmalsvektoren im Bereich unkritischer Verkehrssituationen, also im Bereich schwacher Nachfrage, da diese Zustände häufiger auftreten. Untypische Überlastsituationen sind per Definition selten. Daher werden sich die k nächsten Nachbarn nicht gleichmäßig um das aktuelle Muster verteilen. Dies ist ein Indikator für eine unzuverlässige Prognose. L_X kann klein sein, auch wenn die absoluten Distanzen zwischen aktuellem Muster und den historischen Instanzen groß sind, alleine die Lage ist wichtig.

Distanz der k ähnlichsten $X_{\text{historisch}}$ zu X_{aktuell}

Neben der Lage L_X ist auch die Distanz S_X der zur Prognose verwendeten Instanzen zum aktuellen Muster für die Prognosezuverlässigkeit von Belang. Eine Prognose ist zuverlässiger, wenn die Distanzen des aktuellen Musters X_{aktuell} zu den k verwendeten Distanzen $X_{\text{historisch}}$ klein ist. S_X wird berechnet zu:

$$S_X = \text{Median}\left(X_i^{\text{historisch}} - X_{\text{aktuell}}\right)_{i=\text{Instanz } 1 \dots k}$$

Im Gegensatz zu L_x werden hier die absoluten Distanzen zwischen den Mustern betrachtet. S_x kann entsprechend auch bei perfekt zentraler Lage des aktuellen Musters innerhalb der historischen Instanzen groß werden, wenn die Abstände der Instanzen groß sind.

Verteilung der k ähnlichsten $y_{historisch}$

Neben der Lage und der Ähnlichkeit der $X_{aktuell}$ ist die Eindeutigkeit der zu den k ähnlichsten Instanzen gehörenden abhängigen Variablen $y_{historisch}$ eine wichtige Kenngröße zur Beurteilung der Prognosezuverlässigkeit. Der Erwartungswert y' wird als gewichteter Mittelwert der k ähnlichsten $y_{historisch, i}$ berechnet. Analog dazu wird die gewichtete Varianz σ_y der Stichprobe $y_{historisch}$ berechnet zu:

$$\sigma_y = \frac{\sum_{i=1}^k (w_i \cdot (y_{historisch, i} - y')^2)}{\sum_{i=1}^k w_i}$$

Niedrige Werte für σ_y bedeuten, dass die zu den k ähnlichsten Instanzen gehörenden y_i sehr ähnlich sind. Auf Verkehrsmuster bezogen bedeutet dies, dass die k ähnlichsten Verkehrsmuster der unabhängigen Variablen $X_{historisch}$ alle zu sehr ähnlichen Schätzungen y' führen und von einer hohen Schätzsicherheit ausgegangen werden kann.

Beispiel zur Wirkungsweise der Indikatoren

Zur Erläuterung der drei Zuverlässigkeitsindikatoren wird ein v - q -Diagramm herangezogen. Es wird versucht, die Verkehrsstärke q aus der Geschwindigkeit v mit einem k -Nächste-Nachbarn ($k=51$) Verfahren zu schätzen und umgekehrt. Dabei gehen alle Nachbarn gleich gewichtet bei der lokalen Approximation ein. Als Datenbasis dienen die Minutenwerte eines Tages. Das Ergebnis ist in Abb. 4.13 dargestellt.

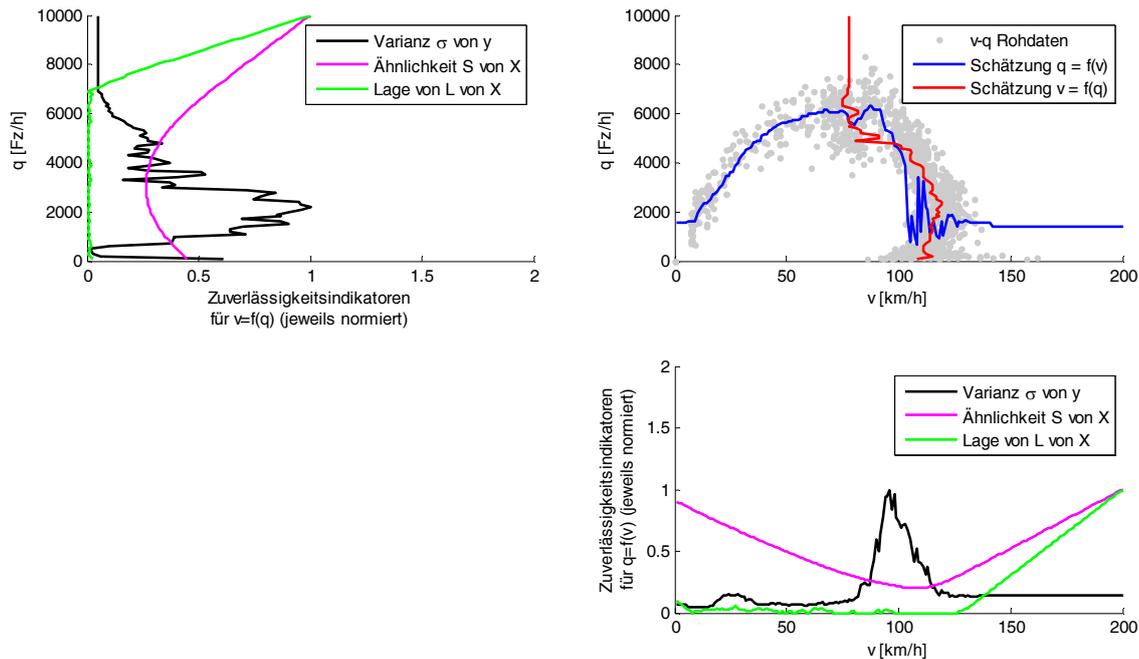


Abb. 4.13 Schätzung von v aus q und umgekehrt für $k = 51$, Daten der A9 München – Nürnberg km 518,530 (AS Garching-Nord), 25. Februar 2006

Wie zu erwarten, funktioniert die Schätzung nur in Teilbereichen zufriedenstellend. Die Schätzung von q aus v ist oben rechts in blau dargestellt. Im Bereich gestauten und gebundenen Verkehrs (etwa $v < 80$ km/h) sieht die Schätzung vernünftig aus. Für 80 km/h $< v < 110$ km/h kann die Verkehrsstärke nicht eindeutig abgeleitet werden, die Verkehrsstärken variieren zwischen null und der Kapazität. Bei zunehmender Geschwindigkeit nehmen die Beobachtungen ab. Die zugehörigen Qualitätsindikatoren sind in der Grafik unten dargestellt. Die Varianz σ_y (schwarz) stellt die Streuung innerhalb der zur Schätzung verwendeten y dar. Im Bereich niedriger Geschwindigkeiten ist σ_y niedrig, der Zusammenhang ist relativ eindeutig. Für 80 km/h $< v < 110$ km/h ist die Varianz aufgrund des uneindeutigen Zusammenhangs größer und fällt wieder für Bereiche hoher Geschwindigkeiten. Die Ähnlichkeit S der aktuellen X zu den historischen Instanzen X ist in magenta dargestellt. Im Bereich hoher Geschwindigkeiten ist der Wert hoch, da die Beobachtungen selten sind und die Nachbarn einen immer größer werdenden Abstand haben. S sinkt mit Geschwindigkeiten um 110 km/h, da hier viele Beobachtungen vorliegen, also zu jedem X_{aktuell} 51 sehr gut passende $X_{\text{historisch}}$ gefunden werden. Im Bereich kleiner Geschwindigkeiten steigt S wieder, da weniger Instanzen vorliegen und das akzeptierte ΔX zum Erhalt der 51 Nachbarn wächst. Die Lage L von X_{aktuell} im Vergleich zu den nächsten Nachbarn ist grün dargestellt. L ist zwischen ~ 10 km/h und ~ 130 km/h niedrig, da das X_{aktuell} hier stets relativ gleichmäßig von den Nachbarn umgeben ist. Über 130 km/h wächst L linear an, da sich X_{aktuell} gleichmäßig von dem fixen Median den 51 höchsten beobachteten Geschwindigkeiten entfernt.

Für die Schätzung von v aus q sehen die Verläufe von S_x und L_x prinzipiell ganz ähnlich aus (in der Grafik oben links). σ_y ist allerdings über weite Bereiche relativ hoch. Zu erklären ist

dies durch den bekannten Effekt, dass allen auftretenden Verkehrsstärken (außer nahe der Kapazität) zwei Verkehrszustände und somit zwei Geschwindigkeiten zugeordnet werden können. Mit steigendem q nimmt σ_y ab.

4.5.2 Schätzung der Wahrscheinlichkeitsdichte des Prognosewertes

Instanzbasiertes Lernen basiert auf der Grundidee, dass die k ähnlichsten historischen Muster identifiziert werden und entsprechend ihrer Ähnlichkeit ein gewichteter Erwartungswert $E(y_i)$ aus den historischen y_i ermittelt wird ($i = 1 \dots k$). Der Erwartungswert ist in vielen Fällen die gewünschte Größe. Neben dem Erwartungswert $E(y_i)$ können weitere statistische Kenngrößen aus den y_i und den zugehörigen Gewichten w_i abgeleitet werden.

Aus Sicht des Zeitpunkts t ist der Prognosewert für den Zeitpunkt $t + \Delta t$ eine Zufallsvariable aus einer zunächst unbekanntem Verteilung. Wenn nun die Wahrscheinlichkeitsdichtefunktion (engl.: *Probability Density Function, PDF*) der Zufallsvariablen bekannt wäre, könnten daraus für den Nutzer interessante statistische Größen abgeleitet werden (beispielsweise Perzentile oder Überschreitungswahrscheinlichkeiten).

Unter der Annahme, dass die k nächsten Nachbarn die wahre Wahrscheinlichkeitsdichtefunktion vollständig repräsentieren (in der Realität zumindest gut approximieren), kann die Zutreffenswahrscheinlichkeit eines y_i einer Instanz i abhängig von ihrer Distanz D_i zum aktuellen Muster abgeschätzt werden. Abb. 4.14 zeigt ein Beispiel zur Analyse der Reisezeiten auf Basis der 13 nächsten Instanzen.

Für jeden Nachbarn i wird die Distanz D_i zum aktuellen Muster ermittelt und anschließend das Gewicht w_i mit dem Gauß-Kern berechnet zu:

$$w_i = e^{-\frac{D_i/D_{i,\min}}{K_w}}$$

Spalten 2 und 3 der Tabelle zeigen die Gewichte und die zugehörigen Reisezeiten. Sortiert man nun die Liste aufsteigend nach der Reisezeit, kann mit der Wahrscheinlichkeitsdichte pE_i (Spalte 9 in der Tabelle, Abbildung oben links) der Instanz i und mit der kumulierten Wahrscheinlichkeitsdichte PE_j (Spalte 10 in der Tabelle, Abbildung oben rechts) die Überschreitungswahrscheinlichkeit der einzelnen Reisezeiten abgeschätzt werden:

$$pE_i = \frac{w_i}{\sum_{i=1}^k w_i} \quad PE_j = \frac{\sum_{i=1}^j w_i}{\sum_{i=1}^k w_i}$$

Neben der Überschreitungswahrscheinlichkeit können auch verschiedene Perzentile (z.B. 5%- und 95%-Perzentile) aus der Verteilung geschätzt werden. Die Wahl von k beeinflusst

das Ergebnis: Je größer k gewählt wird, desto "glatter" wird die Wahrscheinlichkeitsdichtefunktion.

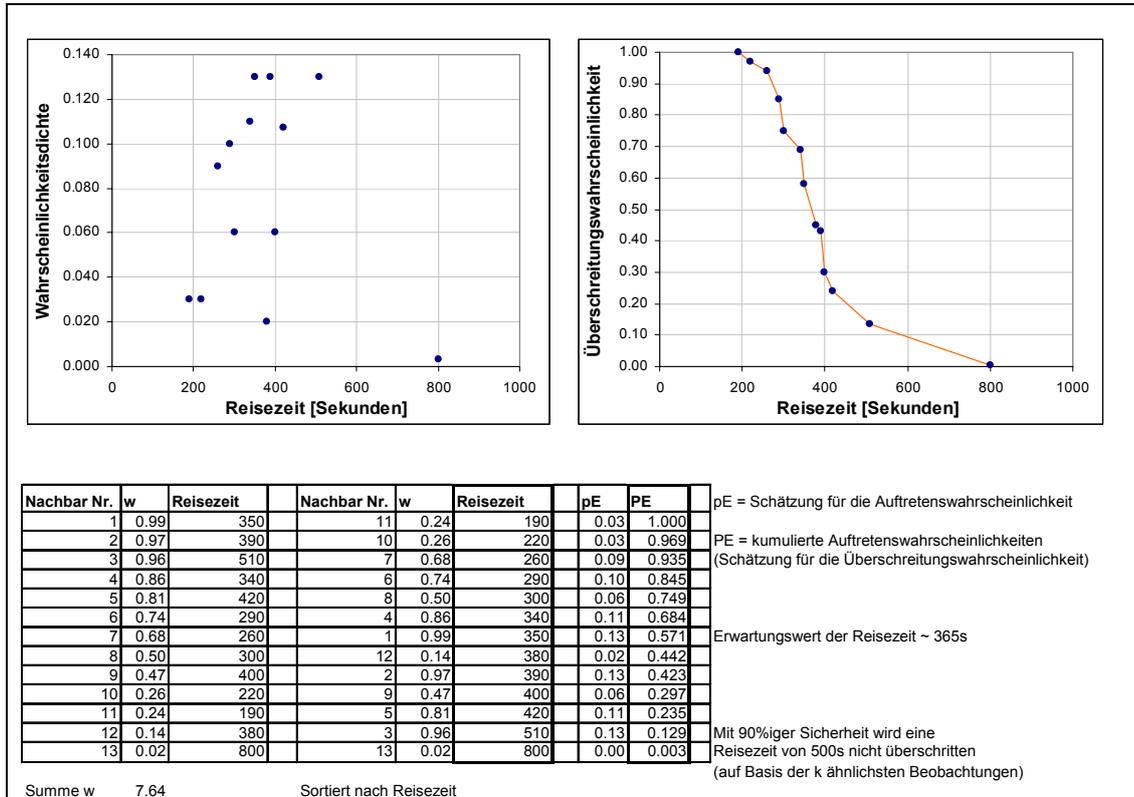


Abb. 4.14 Beispiel zur Abschätzung der Überschreitungswahrscheinlichkeit für die Reisezeit für $k = 13$ Nachbarn

4.6 Gesamtübersicht und Implementierung

4.6.1 Gesamtübersicht über das Verfahren

Abb. 4.15 zeigt einen Gesamtüberblick über die einzelnen Verfahrensschritte. Aus den kontinuierlich erfassten und zufällig erhobenen dynamischen Verkehrsdaten werden lokale Verkehrskenngrößen und Reisezeiten ermittelt. Diese werden unter Berücksichtigung der Geometrie des Straßennetzes in eine räumlich-zeitliche Struktur gebracht, wobei die Verkehrskenngrößen bezüglich des Ortes über den Abstand ds und bezüglich der Zeit über die zeitliche Distanz dt zueinander in Relation gesetzt werden. Die Verkehrskenngrößen werden zusammen mit Kalendervariablen (Tagesklasse, Tageszeit) für den späteren Zugriff fortlaufend archiviert. Die Verkehrskenngrößen der letzten Stunde (in 15-Minutenintervallen) bilden gemeinsam mit den Kalendervariablen das aktuelle Muster. Bei einer Prognoseanfrage (beispielsweise Reisezeit auf Route r im kommenden Zeitintervall $t_{akt} + \Delta t$) werden die für die prognostizierende Kenngröße relevanten Merkmale X_r aus dem Gesamtmuster ausgewählt. Die Relevanz wird anhand der räumlich-zeitlichen Distanzen ermittelt. Für jedes historische

Muster (oder: Instanz) $X_{r,i}$, für das die gewünschte Prognosegröße vorliegt (im Falle von Reisezeiten können das sehr wenige und im Falle lokaler Kenngrößen sehr viele sein), wird die Unähnlichkeit D_i zu dem aktuellen Muster unter Berücksichtigung von ds und dt berechnet. Die k ähnlichsten historischen Instanzen werden zur Prognose herangezogen, wobei sich der Prognosewert y'_r als gewichteter Mittelwert aus den k ähnlichsten $y_{r,i}$ ermittelt wird. Die Gewichtung w_i wiederum basiert auf der Ähnlichkeit des zugehörigen Musters $X_{r,i}$ zum aktuellen Muster. Neben dem Erwartungswert y'_r können Zuverlässigkeitsindikatoren abgeleitet und Verteilungen für y'_r geschätzt werden.

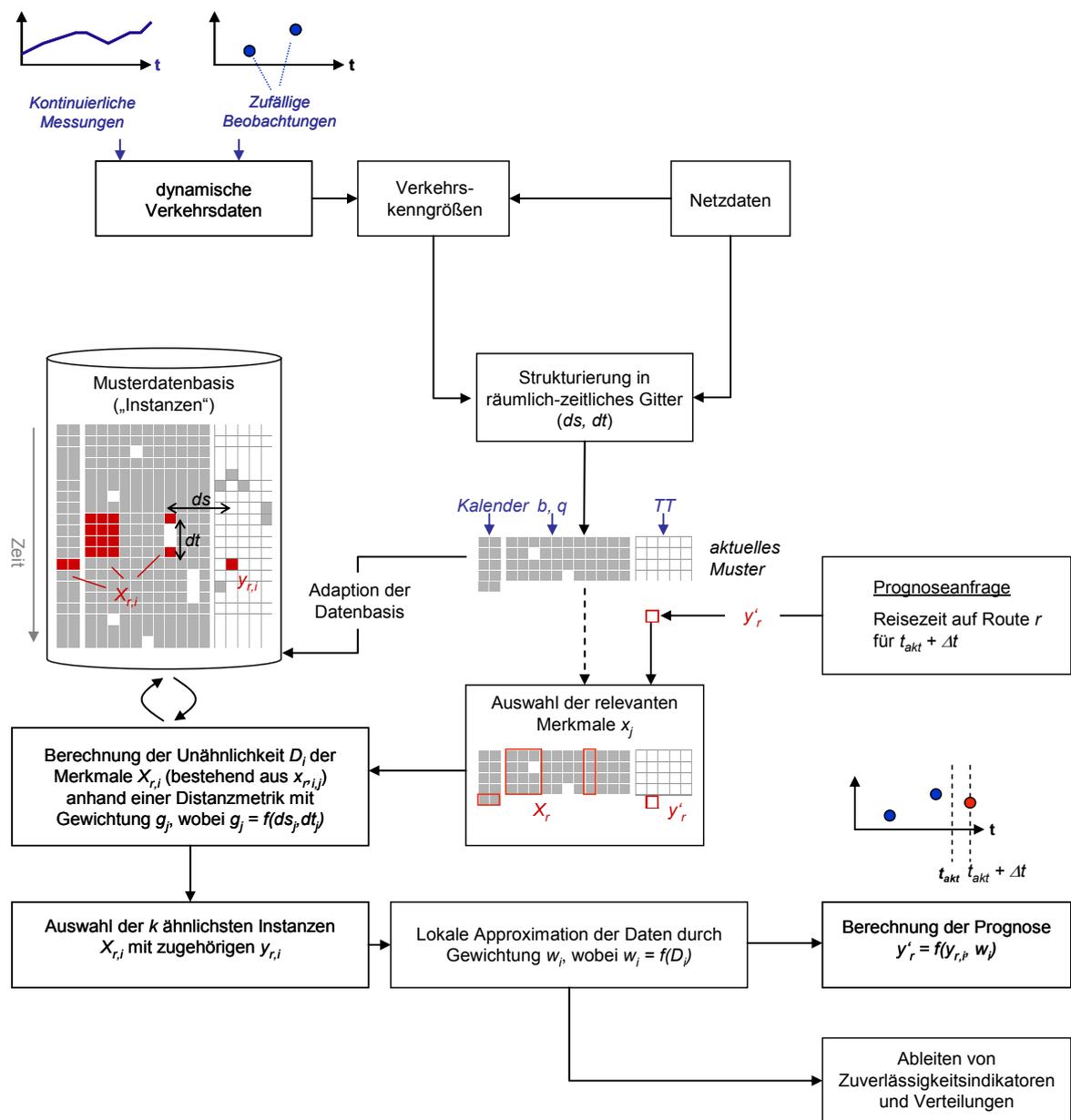


Abb. 4.15 Gesamtübersicht über das Verfahren

4.6.2 Implementierung des Verfahrens

Das Verfahren wurde in der Entwicklungssprache Matlab (Matrix Laboratory) von The Mathworks implementiert. Zum Einsatz im Rahmen des onVLS in Graz wurden die Funktionen mit dem Matlab Compiler als ausführbare Anwendungen (C++) erstellt. Das Programmsystem läuft in den Modi "Online", "Quasionline" und "Offline".

Online bedeutet dabei, dass tatsächlich aktuelle und reale Daten zur Verfügung stehen und verarbeitet werden. Dieser Modus wird für den tatsächlichen Einsatz benötigt. Quasionline bezeichnet den Modus, in dem das Verfahren Daten aus dem Archiv über die Online-Schnittstelle erhält und diese wie im Online Modus verarbeitet. Dieser Modus wird zur Entwicklung und zum Test der Echtzeitfähigkeit des Systems benötigt. Der Offline-Modus wird zur Entwicklung, zum Test und zur Optimierung der eigentlichen Prognosealgorithmen und zur Sensitivitätsanalyse und Optimierung der Modellparameter genutzt (siehe Kapitel 5).

Für die Nutzung der Daten im Prognosemodul werden die Daten im Offline-Modus aus Archivdateien (ASCII Textformate *.txt/*.csv) in mehreren Schritten in den entsprechenden Vorverarbeitungsmodulen (Matlab Funktionen *.m) aufbereitet und dem eigentlichen Prognosemodul in Form von Datenstrukturen (Matlab Datenfelder *.mat) zur Verfügung gestellt (siehe Abb. 4.16).

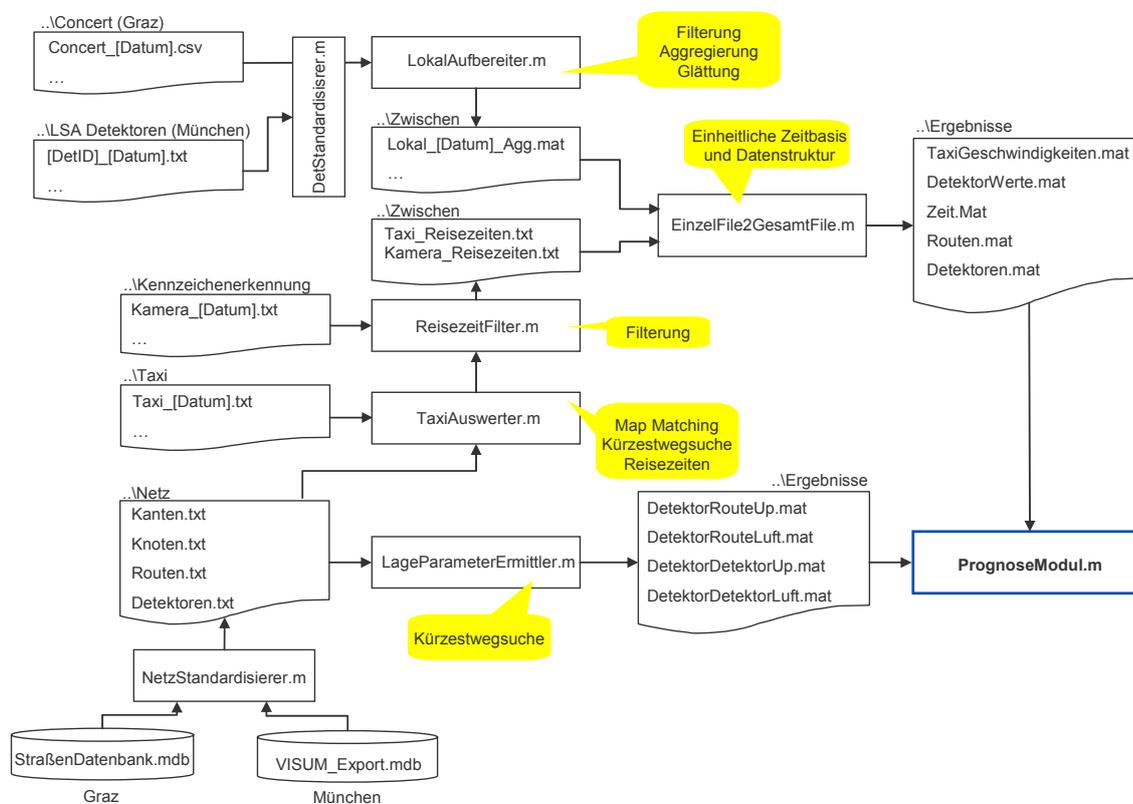


Abb. 4.16 Einbettung des Prognosemoduls in den Datenfluss und das Dateisystem

Als Grundlage werden die in Datenbanken gegebenen Netzdaten in ein einheitliches Netzformat transformiert. Bei den Ausgangsdaten handelt es sich in Graz mit der sogenannten Straßendatenbank um ein spezielles Format und in München um den Standardexport aus einem VISUM (ptv AG) Verkehrsnetz. Auf Basis des standardisierten Netzes werden die Fahrzeugpositionsmeldungen zu Reisezeiten verarbeitet und die Lageparameter der betrachteten Routen und Messstellen für den Mustererkennungsprozess ermittelt. Die aus den Fahrzeugpositionsmeldungen (Graz) und mittels Fahrzeugwiedererkennung (München) gewonnenen Reisezeiten werden soweit möglich gefiltert. Ein wichtiger Schritt zur Verwendung im Prognoseverfahren stellt die Erstellung eines einheitlichen Zeitrasters (hier: 15 Minuten) und die Referenzierung sämtlicher dynamischer Daten auf dieses Raster dar. Anhand der nun vorliegenden zeitlichen Struktur und der in den Lageparametern enthaltenen räumlichen Struktur wird der Mustererkennungsprozess zur Prognose auf Basis der ähnlichsten Instanzen mit räumlich-zeitlicher Gewichtung der Verkehrsmerkmale durchgeführt.

Der modulare Aufbau und die Vereinheitlichung über Standardisierungsmodule ermöglicht eine relativ einfache technische Integration möglicher weiterer Netze und Verkehrsdaten. Zum Test und zur Optimierung wurde das vorgestellte Verfahren zur Prognose von verkehrlichen Kenngrößen in Graz und in München eingesetzt.

4.7 Fazit und weiteres Vorgehen

In Kapitel 4 wurde ein Instanzbasiertes Lernverfahren zur Prognose von Verkehrskenngrößen vorgestellt. Dabei dienen jeweils die Verkehrsmuster relevanter Detektoren, sowie die Tageszeit und die Tagesklasse, den zur Ähnlichkeitsbewertung herangezogenen Merkmalsvektor. Die einzelnen Verkehrsmerkmale werden entsprechend ihrer räumlichen und zeitlichen Beziehung zur zu prognostizierenden Größe gewichtet. Darüber hinaus werden die Kalenderinformationen im Vergleich zum Verkehrsmuster abhängig vom Prognosehorizont gewichtet.

Grundsätzlich berechnet das Modell die Prognose y'_{aktuell} aus dem aktuellen Muster X_{aktuell} , den Instanzen $X_{\text{historisch}}$ und den zugehörigen $y_{\text{historisch}}$. Das Modell besitzt in dieser Form sechs Parameter, die in einem Parametervektor Π zusammengefasst werden können. Somit ist die Notation für die Prognosefunktion für einen gegebenen Parametersatz:

$$y'_{\text{aktuell}} = f(X_{\text{aktuell}}, X_{\text{historisch}}, y_{\text{historisch}}, \Pi)$$

Der Parametervektor Π besteht aus dem Gewicht G_V zur Gewichtung der Ähnlichkeit des Verkehrsmusters, den Parametern k_e , dt_{max} und ds_{max} zur Berechnung der Ähnlichkeit des Verkehrsmusters, dem Parameter k der zu berücksichtigenden Nachbarn und dem Parameter K_W zur Spezifikation der lokalen Approximationsfunktion.

$$\Pi = [G_V, k_e, dt_{\text{max}}, ds_{\text{max}}, k, K_W]$$

Die Modellparameter ermöglichen die Justierung des Modells zum Erhalt einer möglichst optimalen Prognose, bedingen aber eine sinnvolle Einstellung. Im folgenden Kapitel wird untersucht, inwieweit die Prognoseergebnisse von der Wahl der Parameter abhängen. Außerdem wird eine Methode zur automatischen Optimierung der Parameter vorgestellt und das Verfahren anhand realer Daten bewertet.

5 Parametersensitivität und Bewertung des Verfahrens

In Kapitel 5 wird zunächst die Sensitivität der Verfahrensergebnisse bezüglich der sechs Verfahrensparameter untersucht. Aufgrund zu erwartender Wechselwirkungen wird statt einer manuellen Kalibrierung der Parameter mit einem Genetischen Algorithmus ein automatisches Verfahren zur Parametereinstellung vorgeschlagen und exemplarisch angewendet. Anschließend wird das Verfahren für verschiedene Szenarien bewertet und mit anderen Verfahrenstypen verglichen. Schließlich werden die in Kapitel 4 beschriebenen Möglichkeiten zur Schätzung der Prognosezuverlässigkeit anhand verschiedener Situationen untersucht.

5.1 Grundlagen

5.1.1 Vorgehen

Aus der vorhandenen Gesamtdatenbasis (Mai bis Dezember 2007) wird eine Trainingsdatenbasis (1. Mai bis 15. November) entnommen (siehe Abb. 5.1).

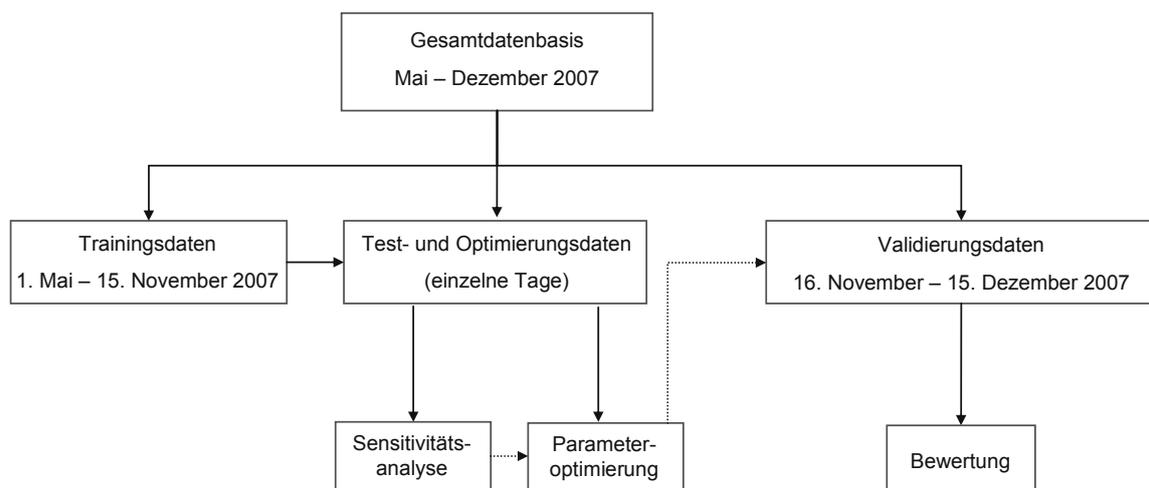


Abb. 5.1 Vorgehen bei Sensitivitätsanalyse, Parameteroptimierung und Bewertung des Verfahrens

Die Trainingsdatenbasis enthält die zur Prognose verwendeten Instanzen und bildet die Grundlage zum Training des in Abschnitt 5.4.2 beschriebenen Künstlichen Neuronalen Netzes. Einzelne Tage der Trainingsdatenbasis werden zur Sensitivitätsanalyse und zur Optimierung der Parameter des Instanzbasierten Lernverfahrens verwendet. Die Sensitivitätsanalyse wird für jeden Modellparameter einzeln zur Identifizierung der Schlüsselparameter des Modells durchgeführt. Zur Optimierung der Parameter wird ein Genetischer Algorithmus (GA) eingesetzt. Der große Vorteil eines GA ist, dass die Parameter gleichzeitig angepasst und über mehrere Generationen in Richtung des optimalen Parametersatzes "geschoben" werden.

Das Verfahren wird anhand von Daten aus dem Zeitraum vom 16. November bis zum 15. Dezember 2007 bewertet und mit anderen Prognosemodellen verglichen.

5.1.2 Methodik der Bewertung

Bei der Bewertung werden die Prognose von Belegungsgraden und die Prognose von Reisezeiten unterschieden.

Mit den gemessenen Belegungsgraden liegen für die Belegungsgradprognose "sichere" Referenzdaten vor. Dabei wird eine funktionierende Sensorik vorausgesetzt. Daher können zur Bewertung der Belegungsgradprognose klassische aggregierte Fehlermaßzahlen zur Beurteilung sinnvoll verwendet werden.

Für die Reisezeiten liegen mit den Taxireisezeiten in Graz Vergleichswerte vor, die nützliche Informationen über die Größenordnung der Reisezeiten liefern, nicht jedoch als 100%ig korrekte Referenz betrachtet werden sollten. Grund ist vor allem, dass mit einer Einzelmessung nur ein kleiner Teil des gesamten Fahrzeugkollektivs der Reisezeit zugrunde liegt. Nicht ganz so gravierend, aber auch gegeben, ist die Problematik bei den Reisezeiten aus Fahrzeugwiedererkennung auf der Route INGOLSTÄDTER, wo Reisezeiten ebenfalls auf einzelnen Fahrzeugen basieren können. Für die Reisezeitprognose werden daher weichere Gütemaße definiert und angewendet.

Bewertung der Belegungsgradprognose (sichere Referenz)

Für die Optimierung und die Bewertung des Verfahrens werden sinnvolle Maßzahlen zur Quantifizierung der Prognosegüte benötigt. Bei Prognoseaufgaben mit vorliegender Referenz werden dazu typischerweise Fehlermaße bezogen auf eine Reihe von PZ realisierten Prognosen verwendet (z.B. $PZ = 96$ für einen Tag bei 15-Minutenintervallen). Gebräuchlich sind und hier verwendet werden die folgenden definierten Maßzahlen.

Beim mittleren absoluten Fehler MAF gehen die Prognosefehler proportional gewichtet ein.

$$MAF = \frac{1}{PZ} \cdot \sum_{p=1}^{PZ} |y_p - y'_p|$$

Beim mittleren quadratischen Fehler MQF werden Fehler quadratisch, also überproportional gewichtet. Ein Fehler der Größe "2" ist also doppelt so schlimm wie zwei Fehler der Größe "1". Die Optimierung bezüglich des MQF führt tendenziell zu einer Vermeidung von sehr großen Fehlern, möglicherweise jedoch auf Kosten von im Mittel größeren Fehlern.

$$MQF = \frac{1}{PZ} \cdot \sum_{p=1}^{PZ} (y_p - y'_p)^2$$

Der maximale Fehler $MAXF$ sollte minimiert werden, wenn seltene sehr große Fehler ein größeres Problem darstellen als viele kleine.

$$MAXF = \max_p |y_p - y'_p|$$

Der mittlere relative Fehler MRF wird auf den jeweiligen Referenzwert bezogen, so dass bei höheren Werten größere absolute Fehler dieselbe Wirkung haben können wie kleinere Fehler bei geringen Prognosewerten.

$$MRF = \frac{1}{PZ} \cdot \sum_{p=1}^{PZ} \frac{|y_p - y'_p|}{y_p}$$

Bei sehr kleinen Prognosewerten y ($\rightarrow 0$) ist der MRF problematisch, da auch sehr kleine absolute Fehler zu sehr großen MRF ($\rightarrow \infty$) führen können. Daher wird stattdessen der mittlere proportionale Fehler MPF verwendet. Bei diesem werden die Fehler ebenfalls auf die Referenzwerte bezogen. Das Problem mit theoretisch unendlich großen Fehlern wird dadurch umgangen, dass der Fehler auf die gesamte Zeitreihe bezogen wird.

$$MPF = \frac{\sum_{p=1}^{PZ} |y_p - y'_p|}{\sum_{p=1}^{PZ} y_p}$$

Bewertung der Reisezeitprognose (unsichere Referenz)

Als Referenz für die mittels fahrzeuggenerierten Daten oder mit Fahrzeugwiedererkennung ermittelte Reisezeit ist die Verwendung der oben genannten Fehlermaße nicht sinnvoll. Die zum Training bzw. zur Prognosebildung verwendeten und als Referenz zur Verfügung stehenden Beobachtungen basieren auf Einzelbeobachtungen und stellen somit zufällige Realisierungen aus den Reisezeitverteilungen aller im betrachteten Zeitintervall vorkommenden Fahrzeuge und nicht notwendigerweise den Erwartungswert dar. Dies gilt vor allem für die Taxireisezeiten, aber auch für die auf der Route INGOLSTÄDTER mit Fahrzeugwiedererkennung erhobenen Reisezeiten (die Stichprobe ist relativ klein, so dass in einem 15 Minuten-Intervall teilweise sehr wenige Fahrzeuge beobachtet wurden). Sie können daher nicht direkt als sichere Referenz für die Prognose eines intervallbezogenen Erwartungswertes verwendet werden. Trotzdem stellen die Einzelfahrzeugreisezeiten wichtige und verwertbare Informationen über den Verkehrszustand auf der jeweiligen Route dar, die eine Beurteilung der Plausibilität der geschätzten und prognostizierten Reisezeiten erlauben.

Eine Möglichkeit zur Verwendung der Beobachtungen als Vergleichswerte ist das unbestrafte Zulassen von Abweichungen beim Vergleich mit der Prognose. Die tolerierten Abwei-

chungen sollten dabei typischen oder maximalen Abweichungen der Einzelmessungen vom jeweiligen Erwartungswert entsprechen. Dadurch wird die Beurteilung weniger scharf, eine fälschliche Bestrafung der Prognose durch einen Einzelwert wird jedoch vermieden. Die absolute Abweichung $e_T(y_p, y'_p)$ mit tolerierter Abweichung Δy^T berechnet sich zu:

$$e_T(y_p, y'_p) = \begin{cases} 0 & \text{wenn } |y_p - y'_p| \leq \Delta y^T \\ |y_p - y'_p| - \Delta y^T & \text{sonst} \end{cases}$$

Die beschriebenen Fehler- oder Abweichungsmaße können unter Berücksichtigung der tolerierten Abweichung berechnet werden, wenn der Term $(y_p - y'_p)$ jeweils durch $e_T(y_p, y'_p)$ ersetzt wird. Aufgrund der einfachen Interpretierbarkeit wird der mittlere absolute Fehler MAE zusätzlich zur "mittleren absoluten Abweichung mit Toleranz", also MAA_T erweitert. Toleriert wird dabei ein Unterschied, der einem zusätzlichen Halt an einer LSA auf der Route entspricht (Annahme: 60 Sekunden).

Allgemein

Neben den Fehlermaßen wird der Grad der linearen Abhängigkeit von Prognose und der Referenz zur Beurteilung der Prognosegüte genutzt. Die lineare Abhängigkeit kann durch das Bestimmtheitsmaß r^2 basierend auf dem empirischen Korrelationskoeffizienten $r_{yy'}$ beschrieben werden.

$$r^2 \text{ mit } r_{yy'} = \frac{\frac{1}{PZ} \cdot \sum_{p=1}^{PZ} (y_p - \bar{y}) \cdot (y'_p - \bar{y}')}{\sqrt{\frac{1}{PZ} \cdot \sum_{p=1}^{PZ} (y_p - \bar{y})^2} \cdot \sqrt{\frac{1}{PZ} \cdot \sum_{p=1}^{PZ} (y'_p - \bar{y}')^2}}$$

Im Gegensatz zu den Fehlermaßen wird hier eine Abweichung toleriert, solange diese systematisch ist. So würde eine konstante Unterschätzung um 20% zu einem Bestimmtheitsmaß von 1, also zu einer 100%igen Abhängigkeit führen, obgleich beispielsweise der *MPF* mit 20% natürlich keine 100%ige Übereinstimmung anzeigen würde. Ein hohes Bestimmtheitsmaß bei gleichzeitig großem Fehler deutet auf eine systematische, also kalibrierbare Abweichung hin. Bei gleichem Fehler ist also das Modell mit dem größeren Bestimmtheitsmaß zu bevorzugen.

Schließlich bildet die grafische Gegenüberstellung der Prognoseergebnisse und der Referenzwerte ein wichtiges Mittel zur Beurteilung der Prognosegüte.

5.2 Modellparametersensitivität

In diesem Abschnitt werden die einzelnen Modellparameter erläutert und der Einfluss der Änderung des Parameters auf den Erwartungswert der Prognose exemplarisch untersucht. Es wird dazu eine Grundeinstellung der Parameter eingestellt und der betrachtete Parameter jeweils innerhalb eines sinnvollen Intervalls variiert.

Parameter	G_v [-]	dt_{max} [15-Min-Intervalle]	ds_{max} [m]	k_e [-]	k [-]	K_w [-]
Grundeinstellung	1	4	5000	2	51	0.1

Tab. 5.1 Grundeinstellung der Parameter

5.2.1 Analyse der einzelnen Parameter

Parameter ds_{max} , dt_{max} und k_e der Distanzmetrik D_v zwischen zwei Verkehrsmustern

Der Bestimmung der Ähnlichkeit zweier Verkehrsmuster kommt eine entscheidende Bedeutung zu. Dies gilt insbesondere bei kleinen Prognosehorizonten, da dann der Einfluss des aktuellen Verkehrsmusters relativ groß im Vergleich zum kalenderbasierten Erwartungswert ist.

Die einstellbaren Parameter zur Bestimmung der Distanz D_v zwischen zwei Verkehrsmustern sind ds_{max} zur Bestimmung des Einflusses der umliegenden Detektoren als Funktion des Abstandes (je niedriger, desto weniger nimmt der Einfluss der Messstellen mit der Distanz ab), dt_{max} zur Bestimmung des Einflusses vergangener Messwerte als Funktion des zeitlichen Abstandes (je niedriger, desto weniger nimmt der Einfluss der Messwerte mit der zeitlichen Distanz ab) und k_e zur Bestimmung des Einflusses größerer Abweichungen (je größer, desto "überproportionaler" werden große Messwertdifferenzen gewichtet).

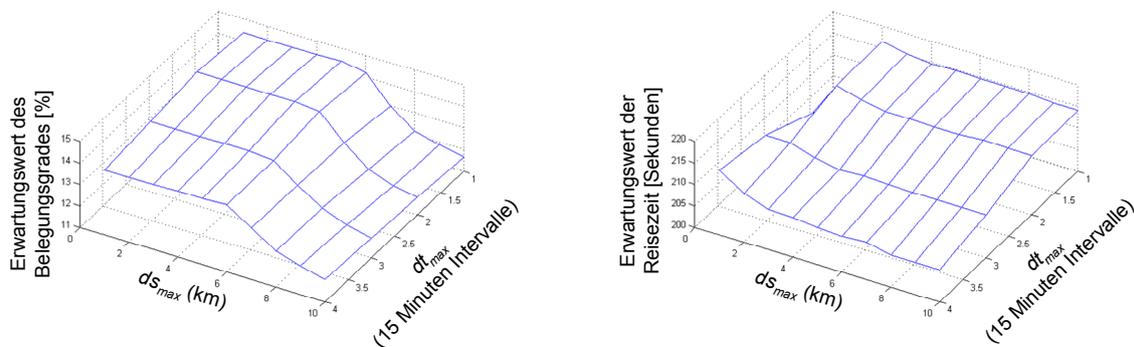


Abb. 5.2 Darstellung der Abhängigkeit des Erwartungswertes von den Parametern ds_{max} und dt_{max} für die Schätzung des Belegungsgrades (links) und der Reisezeit (rechts)

Parameter G_V der Gesamtdistanz D

Der Parameter G_V bestimmt das Verhältnis zwischen kalenderbasierten Prognoseanteilen und dem Einfluss des aktuellen Verkehrsmusters. Die Distanz zweier Verkehrsmuster D_V wird mit dem Faktor G_V gewichtet, wobei G_V abhängig vom Prognosehorizont definiert ist. Zugrunde liegt die Annahme, dass das Verkehrsmuster für kurze Prognosehorizonte (z.B. $\Delta t = 0$ oder $\Delta t = 1 \hat{=} t_{Prog} = 15$ Minuten) im Vergleich zu den Kalendermerkmalen relevanter ist als für große Prognosehorizonte (wie z.B. $\Delta t = 4 \hat{=} t_{Prog} = 60$ Minuten). Es wird angenommen, dass der Einfluss der Verkehrssituation linear mit dem Prognosehorizont t_{Prog} abnimmt.

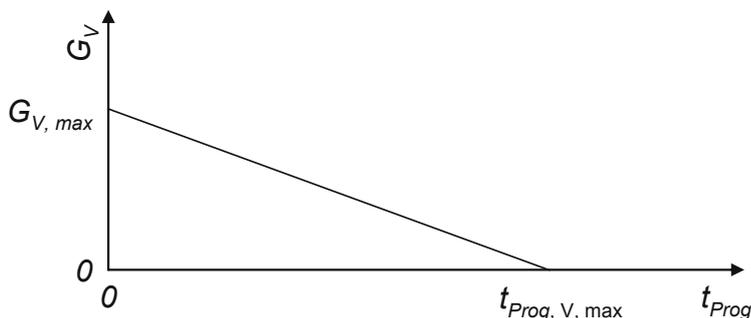


Abb. 5.3 Abhängigkeit von G_V vom Prognosehorizont t_{Prog}

Anzahl der Nachbarn k und Bandbreite K_w des Gauß-Kerns

Der Parameter k ist ein sehr wesentlicher Parameter des Verfahrens. Je größer k gewählt wird, desto konservativer (geglätteter) wird die Prognose, da mehrere Nachbarn berücksichtigt werden. Je kleiner k gewählt wird, desto "spezifischer", aber auch abhängiger von stochastischen Einflüssen wird die Prognose.

Die Wahl von k war und ist Gegenstand zahlreicher Untersuchungen und spielt eine wichtige Rolle bei jeder Applikation von kNN Schätzern. Einige Erkenntnisse aus der Recherche sind die Folgenden:

- Es gibt keine allgemeine Regel zur Wahl von k
- k muss für jeden Anwendungsfall angepasst werden
- $k = 1$ ist häufig die beste Wahl
- k sollte $\leq \sqrt{n}$ ($n =$ Anzahl aller Instanzen des Archivs) gewählt werden, wobei \sqrt{n} häufig die beste Wahl ist (siehe beispielsweise JÖNSSON UND WOHLIN [2004])

Es lässt sich daraus ableiten, dass \sqrt{n} als obere Schranke verwendet werden sollte. Die exakte Einstellung ist Teil eines Kalibrierprozesses.

Abb. 5.4 zeigt beispielhaft wie sich die Wahl von k auf den Erwartungswert auswirkt. Da die Instanzen entsprechend ihrer Distanz gewichtet bei der Berechnung des Erwartungswertes eingehen, konvergiert der Erwartungswert mit steigendem k gegen einen Wert (der Einfluss ist ab einer bestimmten Distanz gleich null). Ab welcher Instanz der Einfluss gegen 0 geht ist abhängig von der Distanz und dem gewählten K_w . Der Parameter K_w des Gauß-Kerns bestimmt die Gewichtung der k nächsten Nachbarn bei der Berechnung des Erwartungswertes. Er dient vor allem der Einstellung, ob der Erwartungswert eher konservativ geschätzt werden soll oder ob nur wenige, zum aktuellen Muster sehr nah liegende Instanzen eingehen sollen.

Wird K_w groß gewählt, geht der negative Exponent des Gauß-Kerns gegen 0 und das Gewicht relativ unabhängig von der Distanz gegen 1. Die k nächsten Nachbarn gehen somit praktisch ungewichtet bei der Berechnung des Erwartungswertes ein (arithmetischer Mittelwert). Wird K_w klein gewählt, wird die Distanz "verstärkt". Da sie negativ im Exponenten des Gauß-Kerns steht, führen größere Distanzen zur überproportionalen Abnahme des Gewichts w . Folglich werden bei der Berechnung des Erwartungswertes ähnliche Instanzen sehr stark gewichtet. In der Anwendung kann das praktisch zu einer Reduzierung der berücksichtigten Nachbarn k führen, da nur ein Teil ein Gewicht $w \gg 0$ zugewiesen bekommt.

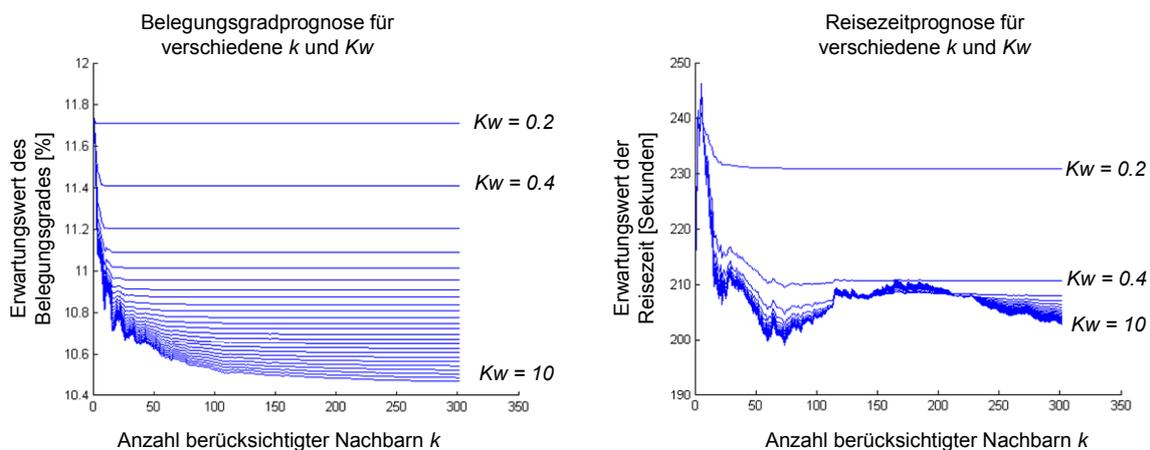


Abb. 5.4 Darstellung des Einflusses von k auf den Erwartungswert für vier Zeitpunkte t_1 bis t_4 für die Belegungsgradprognose (links) und die Reisezeitprognose (rechts)

Abb. 5.5 zeigt das Gewicht der Instanzen als Funktion der Distanz der Instanz (links) und als Funktion des Rangs der Instanz (rechts). Es ist zu erkennen, dass das Gewicht für große K_w (50, 100) praktisch konstant ist. Für kleine K_w strebt w schnell gegen 0, so dass beispielsweise für $K_w = 0.5$ praktisch nicht die ausgewählten 51 Nachbarn sondern nur die 10 ähnlichsten Nachbarn bei der Berechnung der Prognose eingehen (rechtes Bild).

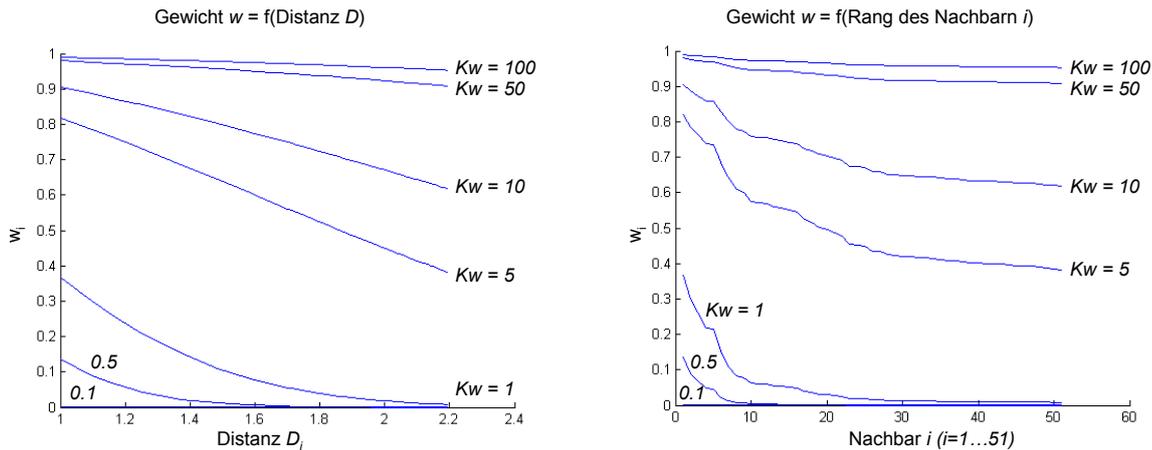


Abb. 5.5 Darstellung des Einflusses des Parameters K_w bei der Gewichtung der Instanzen als Funktion der Distanz der Instanz zum aktuellen Muster (links) und als Funktion des Rangs des Nachbarn (rechts)

Tendenziell wird ein kleines K_w dann günstig sein, wenn tatsächlich nur einige wenige Instanzen eine große Ähnlichkeit aufweisen und für die Prognose relevant sind. Ein großes K_w wäre bei stark verrauschten Werten der historischen Beobachtungen sinnvoll, da hier eine starke Gewichtung eines Wertes zu verrauschten Ergebnissen führen könnte.

5.2.2 Maßzahlen zur Beurteilung der Sensitivität

Die Sensitivität wird exemplarisch für sechs Einzelprognosen, fünf Prognosehorizonte und zwei Szenarien (einmal Belegungsgrad und einmal Reisezeit) untersucht. Die Sensitivität des Verfahrens bezüglich der Parameter zeigt sich in der Streuung des Prognosewertes y' bei unterschiedlichen Parametereinstellungen.

Als Streuungsmaße werden für jeden Parameter P und jeden Prognosehorizont t_{prog} die maximale Spannweite $R(P, t_{prog})_{max}$ und der Median der Spannweiten $\bar{R}(P, t_{prog})_{med}$ der Prognosen ermittelt. Die Spannweite R berechnet sich jeweils als Distanz zwischen dem größten und dem kleinsten Prognosewert über alle realisierten Prognosen PZ :

$$R(P, t_{prog})_{max} = \max_{PZ} \{ y'(P, PZ, t_{prog})_{max} - y'(P, PZ, t_{prog})_{min} \}$$

$$R(P, t_{prog})_{med} = \text{median}_{PZ} (y'(P, PZ, t_{prog})_{max} - y'(P, PZ, t_{prog})_{min})$$

Zusätzlich zu den absoluten Werten werden das Maximum und der Median der Spannweiten bezogen auf den Median der jeweiligen Prognosewerte zur relativen Spannweite berechnet:

$$R(P, t_{prog})_{max,rel} = \max_{PZ} \left\{ \frac{y'(P, PZ, t_{prog})_{max} - y'(P, PZ, t_{prog})_{min}}{\text{median}(y'(P, PZ, t_{prog}))} \right\}$$

$$R(P, t_{prog})_{med,rel} = \underset{PZ}{\text{median}} \left(\frac{y'(P, PZ, t_{prog})_{\max} - y'(P, PZ, t_{prog})_{\min}}{\text{median}(y'(P, PZ, t_{prog}))} \right)$$

Die Spannweite beschreibt, wie groß die Streuung für die gegebenen Randbedingungen durch Variation eines bestimmten Parameters werden kann. Damit enthält sie auch potentielle Ausreißer und zeigt, inwieweit die Variation eines bestimmten Parameters das Ergebnis beeinflussen kann. Im negativen Sinne kann die maximale Spannweite als "Worst Case"-Abschätzung der Parameterabhängigkeit interpretiert werden. Im positiven Sinne zeigt sie, inwieweit das Modell anhand eines bestimmten Parameters kalibriert und somit auch optimiert werden kann. Die mittlere proportionale Streuung stellt den typischen relativen Einfluss bezogen auf den Prognosewert selbst dar. Damit ist auch ein Vergleich der Parameterabhängigkeit verschiedener abhängiger Variablen möglich.

5.2.3 Vergleich der Sensitivitäten

In Tab. 5.2 und Tab. 5.3 sind die Streuungen der Prognose für die sechs Parameter jeweils für fünf unterschiedliche Prognosehorizonte dargestellt. Tab. 5.2 zeigt die absoluten Spannweiten, Tab. 5.3 enthält die relativen Spannweiten der Prognosewerte.

Die Gewichtung des Verkehrsmusters G_v hat sowohl bei den betrachteten Belegungs-, als auch bei den Reisezeitprognosen einen relativ großen Einfluss auf den Prognosewert. Dies zeigt, dass eine Prognose basierend auf Kalendermerkmalen und basierend auf Verkehrsmustern prinzipiell unterschiedliche Ergebnisse zur Folge hat. Der Parameter k_e zur überproportionalen Gewichtung großer Differenzen zwischen einzelnen Verkehrsmerkmalen hat einen geringen Einfluss auf das Prognoseergebnis. Zu interpretieren ist dies mit der Tatsache, dass die betrachteten Verkehrsmuster entweder insgesamt ähnlich oder unähnlich sind und nicht durch einzelne Werte dominiert werden. Der Einfluss von k_e hängt vermutlich auch von dem verwendeten Aggregierungsintervall und dem Glättungsparameter ab. Ebenfalls einen geringen Einfluss hat die Wahl von dt_{max} , also das Gewicht mit dem Messwerte aus vergangenen Intervallen (maximal 60 Minuten) im Vergleich zu den aktuellen Messwerten gewichtet werden. Das legt den Schluss nahe, dass der aktuelle Wert ausreicht, um die Situation zu beschreiben (durch Aggregation und Glättung enthält er ohnehin Informationen aus der Vergangenheit). Andererseits kann es nicht schaden, die Werte der letzten Stunde zu berücksichtigen, da möglicherweise Muster auftreten werden (die in den untersuchten Szenarien nicht enthalten waren), in denen Messwerte der letzten 15 oder 30 Minuten einen beschreibenden Charakter haben. Der Parameter ds_{max} hat einen relativ großen Einfluss auf das Prognoseergebnis, insbesondere bei der Prognose des lokalen Belegungsgrades. Das bedeutet, dass die Information, die in den Zeitreihen unterschiedlicher Detektoren enthalten ist, voneinander abweicht und es eine Rolle spielt, welche Detektoren mit welcher Gewichtung bei der Berechnung der Ähnlichkeit zweier Verkehrsmuster eingehen.

Parameter Wertebereich; Schrittweite Szenario	G_v 0-3;0.2		k_e 1-4;1		dt_{max} 1-4;1		ds_{max} 0.5-10;1		k 1-301;4		K_w 0.001-10;-	
Lokale Belegung (EXERZIERPLATZ) Median Maximum der Streuung [%]												
$t_{Prog} = 0$ Minuten	1.28	1.93	0.19	0.40	0.34	0.47	1.80	2.17	1.22	2.25	0.48	0.91
$t_{Prog} = 15$ Minuten	1.23	1.62	0.15	0.29	0.18	0.35	1.57	2.08	1.10	1.94	0.25	0.68
$t_{Prog} = 30$ Minuten	1.08	1.62	0.06	0.37	0.10	0.25	1.33	1.93	1.04	1.75	0.31	0.50
$t_{Prog} = 45$ Minuten	1.05	1.25	0.18	0.47	0.24	0.76	1.44	1.93	1.17	1.38	0.25	0.47
$t_{Prog} = 60$ Minuten	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.80	6.74	0.00	0.00
Reisezeit Taxi (OPERNRING) Median Maximum der Streuung [Sekunden]												
$t_{Prog} = 0$ Minuten	28	98	17	111	7	17	16	73	100	355	63	289
$t_{Prog} = 15$ Minuten	35	54	10	20	11	15	20	60	103	244	55	137
$t_{Prog} = 30$ Minuten	34	44	12	17	4	7	27	48	67	95	10	38
$t_{Prog} = 45$ Minuten	12	40	9	20	6	9	15	31	79	206	4	5
$t_{Prog} = 60$ Minuten	0	0	0	0	0	0	0	0	114	114	0	0

Tab. 5.2 Absolute Streuungen der Prognose je Parameter und Situation

Parameter Wertebereich; Schrittweite Szenario	G_v 0-3;0.2		k_e 1-4;1		dt_{max} 1-4;1		ds_{max} 0.5-10;1		k 1-301;4		K_w 0.001-10;-	
Lokale Belegung (EXERZIERPLATZ) Median Maximum der Streuung bezogen auf den Prognosewert [%]												
$t_{Prog} = 0$ Minuten	10	16	2	4	3	3	16	19	11	18	4	8
$t_{Prog} = 15$ Minuten	11	15	1	3	2	3	14	19	12	17	2	6
$t_{Prog} = 30$ Minuten	11	17	1	4	1	2	14	20	11	16	3	5
$t_{Prog} = 45$ Minuten	10	14	2	5	3	8	16	21	14	14	3	5
$t_{Prog} = 60$ Minuten	0	0	0	0	0	0	0	0	52	78	0	0
Reisezeit Taxi (OPERNRING) Median Maximum der Streuung bezogen auf den Prognosewert [%]												
$t_{Prog} = 0$ Minuten	12	38	7	38	3	7	6	27	45	164	29	124
$t_{Prog} = 15$ Minuten	15	24	4	8	4	6	9	26	48	107	24	55
$t_{Prog} = 30$ Minuten	14	18	5	7	2	3	11	22	31	44	4	17
$t_{Prog} = 45$ Minuten	6	17	4	9	3	4	7	16	38	97	2	2
$t_{Prog} = 60$ Minuten	0	0	0	0	0	0	0	0	54	54	0	0

Tab. 5.3 Relative Streuungen der Prognose je Parameter und Situation

Die Anzahl der zu berücksichtigenden Nachbarn k stellt einen Schlüsselparameter des Instanzbasierten Lernens dar. Umso überraschender ist das Ergebnis, dass der Erwartungswert der Prognose für den Belegungsgrad bei bis zu 45 Minuten (also solange das Verkehrsmuster berücksichtigt wird) keine große Rolle spielt. Das Ergebnis lässt vermuten, dass die k ähnlichsten Muster (bis zu $k = 301$) sehr ähnlich sind. Für den Belegungsgrad steht ein deutlich größeres Datenarchiv zur Verfügung als für die Reisezeit aus Taxipositionsmeldungen (18651 im Vergleich zu 4259 Beobachtungen für den vorliegenden Fall, 1.5.2007 bis 30.11.2007), da die Belegung bei funktionierenden Detektoren und funktionie-

render Weiterverarbeitung in jedem Zeitintervall erhoben wird. Für einen Prognosehorizont von 60 Minuten hat der Parameter k einen großen Einfluss, da hier ausschließlich Kalendermerkmale verwendet werden. D.h. für $k=1$ wird unter allen jemals für ein bestimmtes Zeitintervall und eine bestimmte Tagesklasse erhobenen Belegungsgraden zufällig eine Beobachtung ausgewählt. Diese kann natürlich vom typischen Wert deutlich abweichen. Bei der Prognose von Reisezeiten hat der Parameter k einen großen Einfluss für alle Prognosehorizonte. Dies ist zum einen mit der kleineren Datenbasis zu erklären (für die Situationen liegen jeweils weniger Beobachtungen vor). Zum anderen spielt sicherlich auch die Tatsache eine Rolle, dass die Reisezeiten in einer bezüglich der Verkehrsmuster der umliegenden Detektoren quasi identischen Verkehrssituation relativ starke Schwankungen aufweisen. Die Wahl von k ist also von großer Bedeutung bei der Prognose der Reisezeiten. Der Parameter K_w definiert den Einfluss der k berücksichtigten Instanzen entsprechend ihrer Distanz und hat einen großen Einfluss auf das Prognoseergebnis. Für $t_{prog} = 60$ Minuten ist der Einfluss null, da die Instanzen "binär gewichtet" eingehen (1 wenn Tagesklasse und Zeitintervall übereinstimmen, 0 wenn nicht).

Abb. 5.6 und Abb. 5.7 zeigen exemplarische Box-Plots für die Prognose des Belegungsgrades und der Reisezeit für verschiedene Prognosehorizonte bei Variation der Parameter k und K_w .

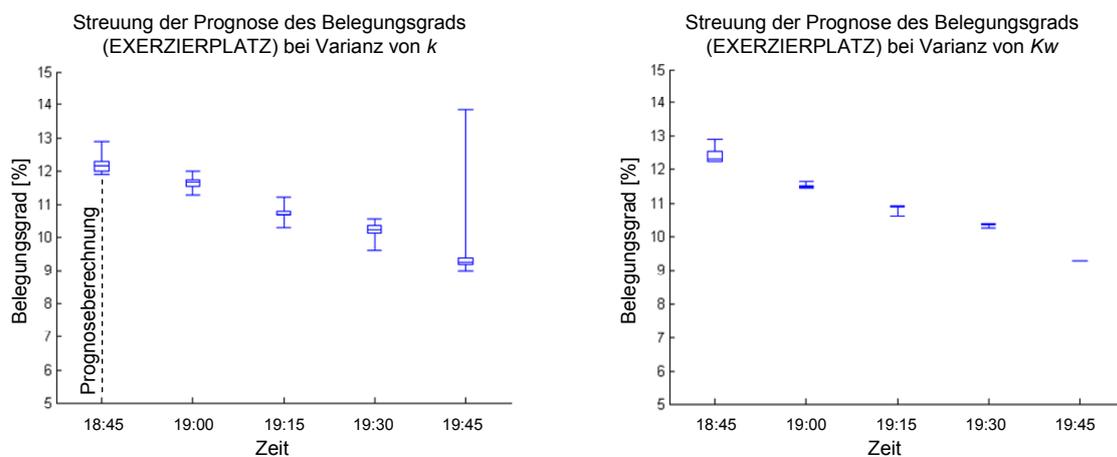


Abb. 5.6 Darstellung der Streuung des Erwartungswertes des Belegungsgrades für verschiedene Prognosehorizonte bei Variation der Parameter k (links) und K_w (rechts)

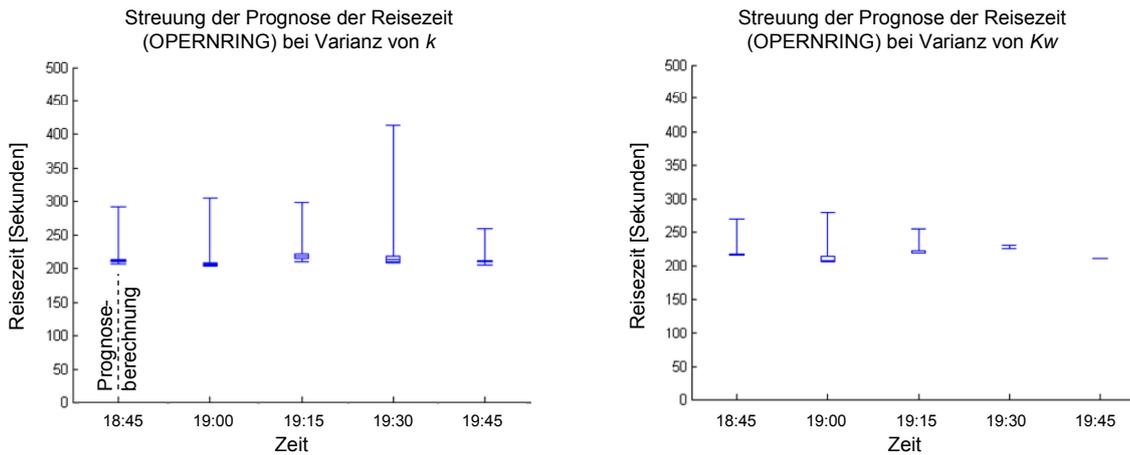


Abb. 5.7 Darstellung der Streuung des Erwartungswertes der Reisezeit für verschiedene Prognosehorizonte bei Variation der Parameter k (links) und K_w (rechts)

5.2.4 Fazit der Sensitivitätsanalyse

Insgesamt ist die Streuung bei Variation der Parameter für die Prognose auf Basis von Taxi-reisezeiten größer als bei den Belegungsgraden. Dies liegt in der deutlich kleineren Datenbasis und der größeren Stochastizität begründet: Schon kleine Änderungen, die zur veränderten Auswahl und Umgewichtung der nächsten Nachbarn führen, können relativ starke Änderungen in der prognostizierten Reisezeit verursachen. Bis auf k_e und dt_{max} haben die Modellparameter einen relativ großen Einfluss auf das Prognoseergebnis, so dass sich ein Ausschluss einer der vier übrigen Parameter bei der Parameterkalibrierung und -optimierung nicht aufdrängt. In der Sensitivitätsanalyse wurden Wechselwirkungen zwischen den einzelnen Parametervariationen ausgeklammert, sind jedoch zu vermuten. Es ist daher anzunehmen, dass die Änderung eines Parameters durch die Änderung anderer Parameter gegebenenfalls kompensiert werden kann.

Es stellt sich daher die Frage des Vorgehens bei der Parametereinstellung. Das optimale Parameterset ist dasjenige, das die Abweichung zwischen Prognose und Referenz minimiert. Dabei kann der Prognosefehler mittels eines der in Abschnitt 5.1.2 beschriebenen Gütemaße quantifiziert werden. Da Wechselwirkungen zu erwarten sind, können die Modellparameter nicht isoliert voneinander betrachtet und optimiert werden. Beispielsweise kann die optimale Einstellung der Anzahl k der zu berücksichtigenden Nachbarn vom gewählten K_w abhängen und umgekehrt. Bei sechs (bzw. vier, wenn k_e und dt_{max} nicht betrachtet werden) zu optimierenden Parametern ist eine manuelle Annäherung an das Optimum zwar möglich, aber sehr aufwändig.

Wünschenswert ist eine automatische, nicht sequentielle Optimierung der Parameter. Bei Existenz einer eindeutigen Zielfunktion (hier: Minimierung der Abweichung) können verschiedene Optimierungsverfahren zur Lösung eingesetzt werden. Gradientenbasierte Suchverfahren benötigen Informationen über die Gradienten zur Minimierung des Fehlers und

gehen bei der Optimierung sequentiell vor. Für das vorliegende Problem erscheinen heuristische Suchverfahren und hier insbesondere Genetische Algorithmen aufgrund ihrer Vorgehensweise (gleichzeitige Optimierung der Parameter), wenig restriktiven Einsatzvoraussetzungen (eine eindeutige Zielfunktion genügt) und der relativ einfachen Applizierbarkeit gut geeignet.

Im folgenden Abschnitt wird der Einsatz eines Genetischen Algorithmus zur Optimierung der Parameter des Instanzbasierten Lernverfahrens beschrieben und exemplarisch angewendet.

5.3 Optimierung der Parameter

5.3.1 Genetischer Algorithmus zur Parameteroptimierung

Genetische Algorithmen (GA) sind stochastische Suchverfahren, die die nächste Lösung durch stochastisches Sampling und regelbasierte Auswahl finden. Dabei ahmen GA den natürlichen Evolutionsprozess nach. Die möglichen Lösungen (hier Parametereinstellungen) werden dabei als Individuen dargestellt. Abb. 5.8 zeigt den Ablauf des GA zur Optimierung der Modellparameter.

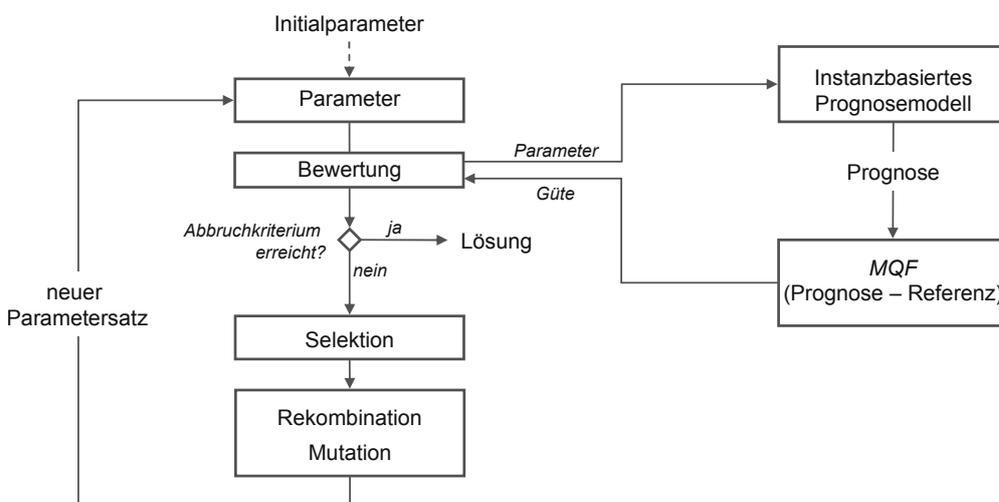


Abb. 5.8 Ablauf der Parameteroptimierung mit einem Genetischen Algorithmus

Zum Beginn des Optimierungsprozesses werden Initiallösungen generiert und anhand einer zu definierenden und eindeutig lösbaren Zielfunktion bewertet. Ist ein Abbruchkriterium erreicht, wird die Lösung verwendet und die Optimierung beendet. Wenn nicht, so werden auf Basis des Selektionsschemas Individuen aus der aktuellen Generation zur Fortpflanzung ausgewählt. Aus den ausgewählten Individuen wird über genetische Operatoren (gebräuchlich sind Rekombination und Mutation) eine neue Generation gebildet und bewertet. Dieser Prozess wird solange wiederholt, bis ein Abbruchkriterium (in der Regel Güte- und Konvergenzkriterium, sowie eine maximale Anzahl an Iterationen) erfüllt ist.

Genetische Algorithmen werden in verschiedenen Bereichen der Verkehrstechnik zur Modell- bzw. Parameteroptimierung eingesetzt. Unter Anderem existieren erfolgreiche Ansätze zur Optimierung von Steuerungsparametern bei der netzweiten Lichtsignalanlagensteuerung (z.B. BRAUN UND WEICHENMEIER [2005], CEYLAN UND BELL [2004]), als Teil hybrider Systeme (z.B. in BOGENBERGER [2001] zur Optimierung eines Fuzzy Logik basierten Zuflussdosierungsverfahrens) und zur Optimierung der Parameter von Verkehrsmodellen (z.B. VORTISCH ET AL. [2008] für Parameter einer Mikrosimulation).

Im Folgenden wird die Spezifikation des GA zur Lösung des Parameteroptimierungsproblems beschrieben. Detailliertere Erläuterungen können in der einschlägigen Literatur gefunden werden (z.B. in GERDES ET AL. [2004], WEICKER [2007]).

Kodierung der Individuen

Die zu optimierenden Modellparameter werden als Individuen reellwertig kodiert. Jedes Gen repräsentiert dabei einen zu variierenden Modellparameter und besitzt einen Wertebereich, innerhalb dessen die optimale Lösung gesucht wird. Jede Generation umfasst 20 Individuen der Form:

Gen	1	2	3	4	5	6
Parameter	G_V	dt_{\max}	ds_{\max}	k_e	k	K_W
Maximum	3	4	10	4	301	10
Minimum	0	1	1	1	1	0.001

Minimum und Maximum stellen dabei den Wertebereich dar, innerhalb dessen das Optimum gesucht wird.

Zielfunktion

Grundlage für die Bewertungsfunktion ist das Prognosemodell. Für vorgegebene Zeitpunkte wird die Prognose berechnet und mit der Referenz verglichen. Als Funktionen zur Bewertung der Lösung wird das in Abschnitt 5.1.2 definierte Fehlermaß MQF , also die mittlere quadratische Abweichung der Einzelprognosen verwendet. Obgleich für die Reisezeitprognose streng genommen nicht von "Fehlerminimierung" (sondern eher von Abweichungsminimierung) gesprochen werden kann, wird das Modell mit dem Ziel einer minimalen Abweichung von Prognose und Taxi- bzw. Fahrzeugwiedererkennungsezeit optimiert. Das stellt unter den gegebenen Randbedingungen ein plausibles Vorgehen dar.

Abbruchkriterien

Das eigentliche Optimierungsziel ist eine möglichst gute Prognose und somit die Minimierung der Zielfunktion, also des MQF . Das erste Abbruchkriterium ist daher:

$$A1: \quad MQF = 0$$

In aller Regel wird dieses Ziel nicht zu erreichen sein. Um zu verhindern, dass der GA unendlich lange nach der optimalen Lösung sucht, wird mit der maximalen Anzahl an Generationen ein zweites Abbruchkriterium definiert.

A2: Maximale Anzahl an Generationen (hier: 14)

Operatoren des GA

Selektion

Bei der Selektion werden Lösungskandidaten aus der aktuellen Generation zur Fortpflanzung per Zufallsprinzip ausgewählt. Grundsätzlich besitzen dabei Lösungskandidaten mit einer hohen Fitness, also entsprechend guten Zielfunktionswerten, eine höhere Fortpflanzungswahrscheinlichkeit. Als Selektionsmethode wird mit der sogenannten *Turnierselektion* eine rangbasierte Selektionsmethode verwendet. Die Auswahlwahrscheinlichkeit hängt dabei ausschließlich von der Rangfolge der Zielfunktionswerte ab. Somit können Dominanzprobleme bei einzelnen extrem viel "fitteren" Individuen – anders als bei Fitnessproportionalen Selektionsverfahren – effektiv vermieden werden. Zur Auswahl werden jeweils n_i Individuen (hier: $n_i = 2$) zufällig gezogen und dasjenige mit der höchsten Fitness zur Fortpflanzung ausgewählt.

Rekombination

Die zur Fortpflanzung selektierten Individuen werden mit einer vorzugebenden Wahrscheinlichkeit rekombiniert. Rekombination bedeutet die Kreuzung von zwei Individuen. Bei reellwertiger Kodierung der Individuen kommen i.d.R. verschiedene Rekombinationsstrategien zum Einsatz. Hier werden die *einfache* und die *arithmetische* Rekombination verwendet. Bei der einfachen Rekombination werden die Gene der Eltern ab einer zufällig ausgewählten "Bruchstelle" getauscht, bei der arithmetischen Rekombination werden die Gene der Kinder als gewichteter Mittelwert der jeweiligen Gene der Eltern ermittelt. Die Gewichtung wird entsprechend einer Zufallszahl zwischen 0 und 1 eingestellt.

Mutation

Die Mutation führt neben der Rekombination eine zusätzliche Zufallskomponente ein. Dabei werden die Gene der Individuen mit einer bestimmten Wahrscheinlichkeit ("Mutationsrate") zur Mutation ausgewählt und verändert. Die Mutation dient zur Generierung einer gewissen Inhomogenität und Divergenz in der Population und verhindert eine frühzeitige Konvergenz bei der Optimierung. Hier wird die *Uniform Mutation* verwendet. Dabei wird ein zufällig ausgewähltes Gen des Individuums verändert.

Es existieren noch eine Reihe weiterer genetischer Operatoren wie die Permutation, die Inversion oder die Verschiebung von Genen (siehe beispielsweise GERDES ET AL. [2004]), die hier aber nicht weiter betrachtet werden.

Implementierung

Der Genetische Algorithmus wurde in Matlab auf Basis der "Genetic Algorithm Optimization Toolbox" (GAOT, von Chris Houck, Jeff Joines, und Mike Kay an der North Carolina State University entwickelt, www.ise.ncsu.edu/mirage/GAToolBox/gaot (Stand: 12.06.1998)) implementiert.

5.3.2 Ergebnisse der Parameteroptimierung

Der GA wurde zur Optimierung der Prognose für sämtliche Detektoren und Routen für jeweils drei Prognosehorizonte eingesetzt (Ergebnisse siehe Tab. 5.4). In Abb. 5.9 ist die Performanz in Form des Zielfunktionswerts als Funktion der Generation dargestellt. Sowohl für den Belegungsgrad, als auch für die Reisezeit ist deutlich zu sehen, dass die Prognosegüte für die Initiallösungen ebenso wie im weiteren Verlauf für die kleineren Prognosehorizonte höher ist. Dies stellt zwar kein überraschendes Ergebnis dar, zeigt jedoch, dass das Verfahren unter günstigeren Bedingungen auch besser arbeitet, was insgesamt plausibel ist.

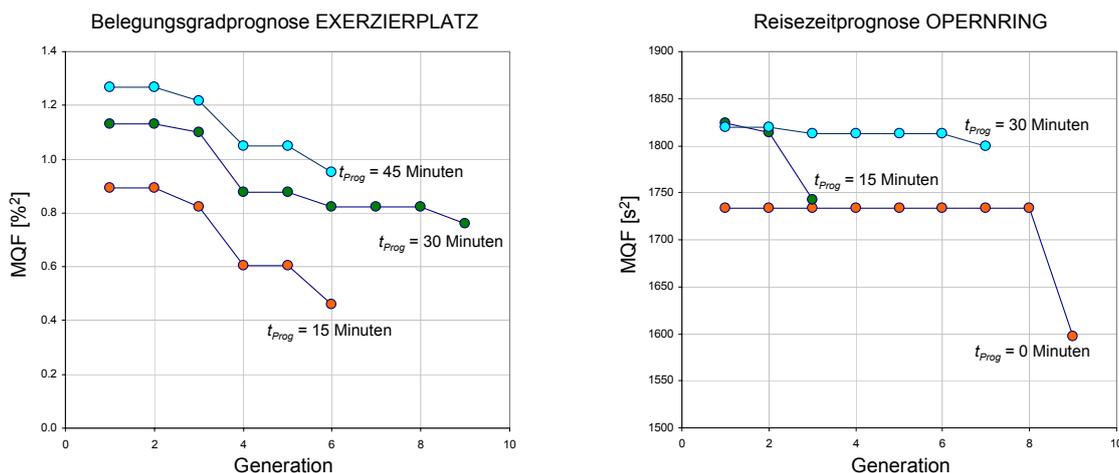


Abb. 5.9 Visualisierung des Zielfunktionswertes MQF bei der Optimierung der Belegungsgradprognose (links) und der Reisezeitprognose (rechts) mit GA

Als Zielfunktion wurde die mittlere quadratische Abweichung zwischen prognostiziertem und gemessenem Belegungsgrad verwendet. Für die Reisezeit wurde als Zielfunktion die mittlere quadratische Abweichung zwischen prognostizierter Reisezeit und Taxireisezeit bzw. der Reisezeit aus Fahrzeugwiedererkennung verwendet. Eine tabellarische Darstellung der Ergebnisse der Optimierungsläufe über die verschiedenen Generationen ist in Anhang 2 zu finden.

In den meisten Szenarien zeigt sich ein hoher Wert von G_V , also eine hohe Gewichtung des Verkehrsmusters relativ zu den Kalendermerkmalen als günstig. Dies gilt interessanterweise auch für große Prognosehorizonte, wobei G_V mit steigendem Prognosehorizont ohnehin linear abnehmend eingestellt ist (siehe Abschnitt 5.2.1), so dass hier die Kalendermerkmale

trotzdem stärker gewichtet eingehen. Häufig bewirken die Kalendermerkmale auch eine Art Vorsortierung (analog zur Vorklassifizierung), wobei innerhalb der kalendarisch ähnlichen Muster die ähnlichsten Verkehrsmuster gesucht werden.

k_e ist eher niedrig zu wählen. D.h., eine überproportionale Gewichtung einzelner starker Abweichungen im Verkehrsmuster ist nicht sinnvoll. Ausnahme bildet hier der OPERNRING, auf dem für alle untersuchten Prognosehorizonte ein hohes k_e günstig ist.

Für das optimale dt_{max} ist keine klare Tendenz zu erkennen, der Parameter ist anscheinend stark lokationsabhängig. Möglicherweise spielen auch die Eigenschaften der bei der Optimierung zugrunde gelegten Datensätze eine Rolle.

Insgesamt zeigen sich für alle Szenarien relativ kleine Werte für ds_{max} als günstig. Das bedeutet, dass die starke Gewichtung entfernter Messstellen tendenziell ungünstig ist. Wenn also Messstellen auf einer betrachteten Route zur Verfügung stehen, kann es besser sein, nur diese zur Prognose zu verwenden und nicht auch entfernte. Dies gilt analog für die Prognose des lokalen Belegungsgrades. Hier sind die wesentlichen Informationen in der zu prognostizierenden Zeitreihe selbst enthalten. Für die Reisezeit auf der INGOLSTÄDTER befinden sich alle verwendeten Detektoren auf der Route selbst, so dass hier eine Gewichtung entsprechend der Entfernung nicht vorgenommen wird.

Die Werte k und K_w beeinflussen sich gegenseitig. Die Berücksichtigung nur weniger Nachbarn k spielt die Bandbreite K_w eine geringere Rolle, während für große Bandbreiten K_w die Anzahl der Nachbarn eine größere Rolle spielt (da sie tendenziell gleich gewichtet berücksichtigt werden) und umgekehrt. Die Wahl ist auch hier lokations- und prognosehorizontabhängig zu treffen. Für die Belegungsgradprognose zeigt sich die Wahl eines hohen K_w für große Prognosehorizonte als günstig, d.h. für größere Horizonte werden die Instanzen tendenziell gleich gewichtet, was zu einer stärkeren Glättung führt. Für die Reisezeitprognose ist diese Tendenz nicht klar erkennbar.

Szenario	Gv	K_e	dt_{max}	ds_{max}	k	K_w	MQF
EXERZIERPLATZ							
$t_{Prog} = 15$ Minuten	9.82	1.75	4.57	2.42	172	1.40	0.33
$t_{Prog} = 30$ Minuten	10	1.51	10	2.47	101	4.97	0.45
$t_{Prog} = 45$ Minuten	5.11	1.74	6.61	2.74	25	9.53	0.57
ERZHERZOG							
$t_{Prog} = 15$ Minuten	10	1	5.30	2.43	56	6.28	0.46
$t_{Prog} = 30$ Minuten	8.15	1	3.52	3.08	48	5.70	0.76
$t_{Prog} = 45$ Minuten	10	1	1	1	101	10	0.95
OPERNRING							
$t_{Prog} = 0$ Minuten	9.02	8.40	1.30	3.70	32	0.19	1598

$t_{Prog} = 15$ Minuten	8.52	4.50	1.00	3.84	63	9.16	1742
$t_{Prog} = 30$ Minuten	9.03	8.60	2.25	1.56	128	5.99	1799
KEPPLER-WEST							
$t_{Prog} = 0$ Minuten	8.74	1.29	2.36	3.17	100	3.45	2583
$t_{Prog} = 15$ Minuten	9.02	8.40	1.30	3.70	32	0.19	2600
$t_{Prog} = 30$ Minuten	9.43	1.02	1	1.07	298	9.10	2736
INGOLSTÄDTER							
$t_{Prog} = 0$ Minuten	8.73	1.29	2.36	-	71	0.35	4099
$t_{Prog} = 15$ Minuten	1	1	1	-	1	10	4440
$t_{Prog} = 30$ Minuten	10	1	6.18	-	290	1	6502

Tab. 5.4: Optimierte Parameter

5.3.3 Fazit der Parameteroptimierung

Der Einsatz von GA zur Parameteroptimierung in dem hier verwendeten Instanzbasierten Lernverfahren ist prinzipiell möglich und führt zu verbesserten Ergebnissen im Vergleich zur gewählten Standardparametereinstellung. Allerdings muss die Optimierung streng genommen für alle Szenarien (Prognosehorizonte, Lokationen) getrennt durchgeführt werden, da eine Übertragbarkeit der optimalen Parameter augenscheinlich nicht möglich ist. Dabei kommt mit der relativ langsamen Ausführungsgeschwindigkeit der größte Nachteil des Instanzbasierten Lernens voll zu Geltung und der Optimierungsprozess kann sehr lange dauern. Für praktische Anwendungen kann der Prozess der Optimierung für alle relevanten Szenarien wiederum mittels Stapelverarbeitung automatisiert werden, so dass damit ausschließlich der Rechner und nicht sein Benutzer belastet wird. Für weitere Optimierungsläufe sollten die oberen bzw. unteren Grenzen derjenigen Parameter, deren Optimum eben einen dieser Grenzwerte angenommen hat, entsprechend angepasst werden (z.B. G_V und K_W).

Bei der Parametereinstellung können neben der Minimierung des Fehlers weitere Kriterien eine Rolle spielen. Beispielsweise kann durch entsprechende Parameterwahl eine mehr oder weniger starke Glättung eingestellt werden.

Im kommenden Kapitel wird das Verfahren für verschiedene Szenarien getestet und im Vergleich zu anderen Verfahren bewertet. Dabei werden die mittels Optimierung erhaltenen Parametersätze eingesetzt. Für Prognosehorizonte, die bei der Optimierung nicht abgedeckt wurden, werden die am besten passenden Parameter (z.B. Nutzung der 15 Minuten-Parameter für die 0-Minuten-Prognose etc.) verwendet.

5.4 Eingesetzte Prognoseverfahren

5.4.1 Instanzbasierte Lernverfahren

Das in Kapitel 4 beschriebene Instanzbasierte Lernverfahren wird in drei unterschiedlichen Implementierungen getestet.

"IL": Die Standardimplementierung prognostiziert den Erwartungswert direkt aus den Trainingsdaten, möglicherweise vorliegende aktuelle Messwerte der Zielgröße werden nicht berücksichtigt. Damit handelt es sich um einen reinen Mustererkennungsprozess und die Voraussetzungen an die Datenverfügbarkeit der Zielgröße sind gering (es genügen zufällige historische Beobachtungen).

"IL mit GL Anpassung" (nur Belegungsgradprognose): Für Prognosehorizonte bis 30 Minuten wird statt des Erwartungswertes eine modifizierte Prognose auf Basis des mit "IL" ermittelten Erwartungswertes und des aktuellen Messwerts mittels der Angleichmethode (Abschnitt 4.4.1) berechnet. Ist dies nicht möglich, etwa weil kein aktueller Messwert vorliegt, wird der ursprünglich ermittelte Erwartungswert verwendet.

"IL ohne GL": Der Mustervergleich wird ohne die Berücksichtigung der Werte des zu prognostizierenden Detektors (Belegungsgrad) bzw. ohne die Berücksichtigung der Detektoren auf der betrachteten Route (Reisezeit) durchgeführt. Folglich ist neben den "echten" Prognosen auch die Schätzung ($t_{Prog} = 0$ Minuten) der Kenngröße eine mögliche Aufgabe. Als Muster zur Identifikation der ähnlichsten Instanzen kommen daher nur die Kenngrößen umliegender Detektoren in Frage.

5.4.2 Benchmarkverfahren

Das neu entwickelte Prognoseverfahren trifft als ein weiteres Mitglied auf eine Familie existierender Verfahren und Ansätze (siehe Kapitel 3) und wird daher mit einigen Verfahren verglichen. Ausgewählt und implementiert wurden drei relativ einfache (kalenderbasierte Prognose, Prognose auf Basis lokaler Muster und Trivialprognose) und mit einem künstlichen Neuronalen Netz ein komplexeres Verfahren.

Einfache kalenderbasierte Prognose

Für große Prognosehorizonte (etwa ≥ 30 Minuten) und in typischen Verkehrssituationen stellen *einfache kalenderbasierte*, also auf Tagesganglinien basierende Erwartungswerte eine wichtige Klasse von Prognoseverfahren dar. Dabei wird der zu prognostizierende Wert als Erwartungswert sämtlicher Beobachtungen eines Tagestyps und eines Zeitintervalls ermittelt. Hierbei können naturgemäß keine aktuellen verkehrlichen Abnormalitäten einfließen.

Kalenderbasierte Prognosen liefern unter typischen Bedingungen üblicherweise gute Ergebnisse.

Lokale Mustererkennung

Die Prognose auf Basis der lokal gemessenen Ganglinie entspricht im Prinzip dem Instanzbasierten Lernen auf lokaler Ebene, also bei ausschließlicher Berücksichtigung der Werte an dem betrachteten Detektor selbst. Dabei wird die letzte Stunde der aktuellen Ganglinie mit historischen Ganglinien-Abschnitten aus demselben Stundenintervall und derselben Tagesklasse verglichen. Die ähnlichste(n) historischen Ganglinien werden zur Prognose herangezogen. Ganglinien anderer Messstellen werden nicht berücksichtigt. Dieses Vorgehen ist häufig der Standard zur Prognose der Nachfrage für Verkehrsmodelle (VORTISCH [2005]) oder dient als Grundlage für weitergehende Mustererkennungsprozesse (VON DER RUHREN [2006]).

Die lokale Mustererkennung wird im Folgenden mit "IL Lokal" bezeichnet. Die lokale Mustererkennung mit Einbezug (Fortschreibung) der aktuellen Ganglinien wird mit "IL Lokal GL" bezeichnet.

Trivialprognose

Bei kontinuierlich vorliegenden Zeitreihen der zu prognostizierenden Größe (i.d.R. bei lokalen Messwerten der Fall) liefert die *Trivialprognose* für kurze Prognosehorizonte (etwa ≤ 15 Minuten) gute Ergebnisse, die von einem komplexeren Verfahren nicht zwingend erreicht werden (siehe beispielsweise WILD [1996]). Bei der Trivialprognose wird der aktuelle Messwert einfach konstant in die Zukunft extrapoliert. Für einen Prognosehorizont dt ergibt sich die Trivialprognose zu:

$$y'(t + dt) = y(t)$$

Die Trivialprognose wird nur für die Prognose des Belegungsgrades angewendet.

MLF (Künstliches Neuronales Netz)

Aufgrund ihrer Fähigkeit, nichtlineare Zusammenhänge abbilden zu können und diese bei Vorliegen eines Trainingsdatensatzes automatisiert zu lernen, sind *Künstliche Neuronale Netze* (engl.: *Artificial Neural Networks ANN*) eine in den letzten Jahren sehr populäre Modellklasse für Prognosefragestellungen (siehe Abschnitt 3.4). Dabei existiert für die Klasse der ANN eine Vielzahl möglicher Netzarten und -architekturen. Eine bei Mustererkennungsproblemen häufig eingesetzte Klasse sind so genannte Multilayer Feedforward (MLF) Netze. MLF können durch den Einsatz effizienter Lernalgorithmen die optimalen Gewichts- und Übertragungsfunktionen für eine gegebene Netzstruktur einstellen. Jedoch bleibt der eigent-

liche Netzentwurf, also die Wahl der Anzahl und der Art der Eingänge (entspricht dem Merkmalsvektor), die Wahl der Anzahl verdeckter Schichten und die Wahl der Anzahl von Neuronen je Schicht dem Entwickler oder wiederum einem externen Optimierungsverfahren überlassen.

Im Gegensatz zu "der" Trivialprognose gibt es also nicht "das" MLF. Auf Basis der Literaturrecherche und einer Analyse der Eingangsdaten werden zwei MLF Topologien entworfen und als Benchmarkverfahren implementiert. Diese besitzen jeweils zwei versteckte Schichten (3/10 Neuronen bzw. 5/15 Neuronen), lineare Transferfunktionen zwischen den versteckten Schichten und der Ausgabeschicht und Sigmoid-Transferfunktionen zwischen Eingabeschicht und erster versteckter Schicht.

Als Eingangsvektoren werden hier dieselben Merkmale wie für das Instanzbasierte Lernverfahren verwendet (siehe Abb. 5.10). So können zwei unterschiedliche Lernmethoden bzw. Mustererkennungsansätze (Instanzbasiert und MLF Training) direkt miteinander verglichen werden. Grob skizziert umfassen Training und Anwendung des MLF folgende Schritte:

Training: 1. Bereinigung der Trainingsdaten; 2. Normierung der Trainingsdaten; 3. Training des Neuronalen Netzes

Anwendung: 1. Bereinigung der Anwendungsdaten; 2. Normieren der Anwendungsdaten; 3. Simulation durch das trainierte Netz; 4. Rücktransformation der normierten Ausgabe in den ursprünglichen Wertebereich

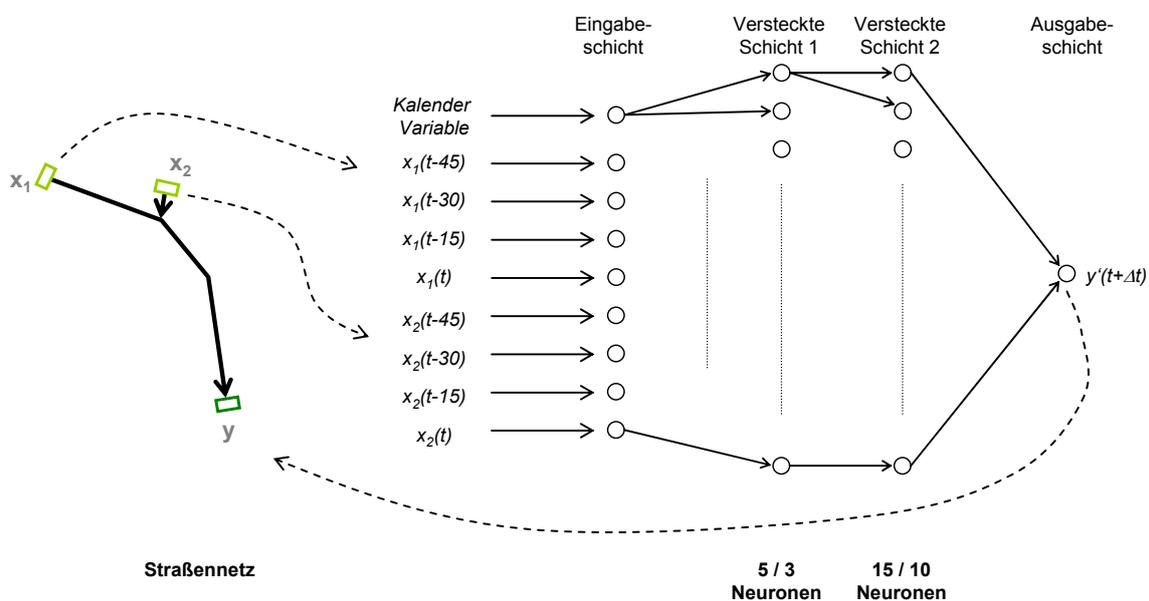


Abb. 5.10 MLF zur Abbildung der Zusammenhänge des Belegungsgrades an einem Detektor (y) durch die Belegungsgrade der vier letzten Zeitintervalle zweier stromaufwärts gelegener Detektoren (x_1 und x_2)

Nach der Auswahl der Netztopologie besteht bei einer definierten Trainingsbasis die Hauptschwierigkeit darin, das Netz so zu trainieren, dass es einerseits den Fehler minimiert und andererseits seine Generalisierungsfähigkeit erhält (vermeiden von "Overfitting"). Um diese beim Training widersprüchlichen Optimierungsziele zu kombinieren, können zwei Techniken eingesetzt werden: (1) "Early Stopping", d.h. das Training wird gestoppt, bevor der Fehler minimiert ist und (2) "Regularisierung" durch Bestrafung von hohen Verbindungsgewichten zwischen den Neuronen. Da "Early Stopping" keine Konvergenz bedingt, wird es hier als Technik zur Vermeidung von Overfitting verwendet.

Abb. 5.11 zeigt das Lernverhalten des MLF (Architektur: 2 versteckte Schichten mit 5 bzw. 15 Neuronen). Bei der Prognose des Belegungsgrades (links) konvergiert die Performanz (Kennwert aus dem mittleren quadratischen Fehler) nach 9 Trainingsepochen für die 15-Minuten-Prognose bzw. nach 6 Epochen für die 45-Minuten-Prognose. Dabei ist die Performanz für die 15-Minuten-Prognose erwartungsgemäß besser. Bei der Optimierung zur Reisezeitprognose konvergiert das Netz schneller (nach 5 bzw. 3 Epochen) und die Performanz ist schlechter. Dies überrascht wiederum nicht, da die Reisezeiten zum einen stärker verrauscht sind und zum anderen der Zusammenhang nicht auf Basis der Größe selbst (Reisezeit aus Belegung im Gegensatz zu Belegung aus Belegung) gelernt wird.

Die getesteten MLF werden mit "MLF 1" (zwei versteckte Schichten mit 3 bzw. 10 Neuronen) und "MLF 2" (zwei versteckte Schichten mit 5 bzw. 15 Neuronen) bezeichnet. Das MLF 1 wird außerdem für den Fall "aktuelle Messwerte der betrachteten Ganglinie sind nicht vorhanden" getestet ("MLF 2 ohne GL").

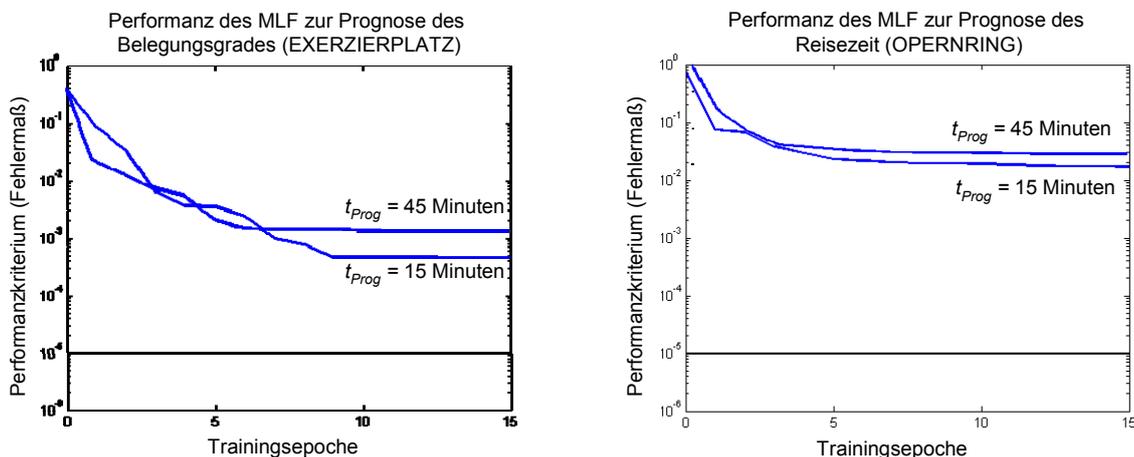


Abb. 5.11 Performanz des MLF als Funktion der Trainingsepochen für die 15-Minuten- und 45-Minuten-Prognose des Belegungsgrades EXERZIERPLATZ (links) und der Reisezeit OPERNRING (rechts)

Die MLF Applikationen wurden unter Nutzung der "Neural Networks Toolbox" von Matlab implementiert.

5.5 Ergebnisse der Belegungsgradprognose

In diesem Abschnitt werden die Ergebnisse der Belegungsgradprognose dargestellt und diskutiert. Teilweise sind die verschiedenen Fehlermaße redundant, werden der Vollständigkeit wegen jedoch trotzdem dargestellt. Der mittlere absolute Fehler und das Korrelationsmaß als Gütemaße sind tabellarisch aufgelistet. Der mittlere proportionale und der mittlere quadratische Fehler sind als Netzgrafiken dargestellt (Darstellung des Prognosefehlers: Je weiter vom Mittelpunkt des Netzes entfernt, desto größer ist der Prognosefehler). In Anhang 3 befinden sich weitere Darstellungen (Zeitreihen und Scatter-Plots).

EXERZIERPLATZ (1.-7. Dezember 2007)

Die Ganglinie am Detektor EXERZIERPLATZ zeigt üblicherweise eine mehr oder weniger deutlich ausgeprägte Morgenspitze (siehe beispielsweise Abb. 5.12), die Zeitreihe zeigt über den Tag gesehen selten stationäres Verhalten

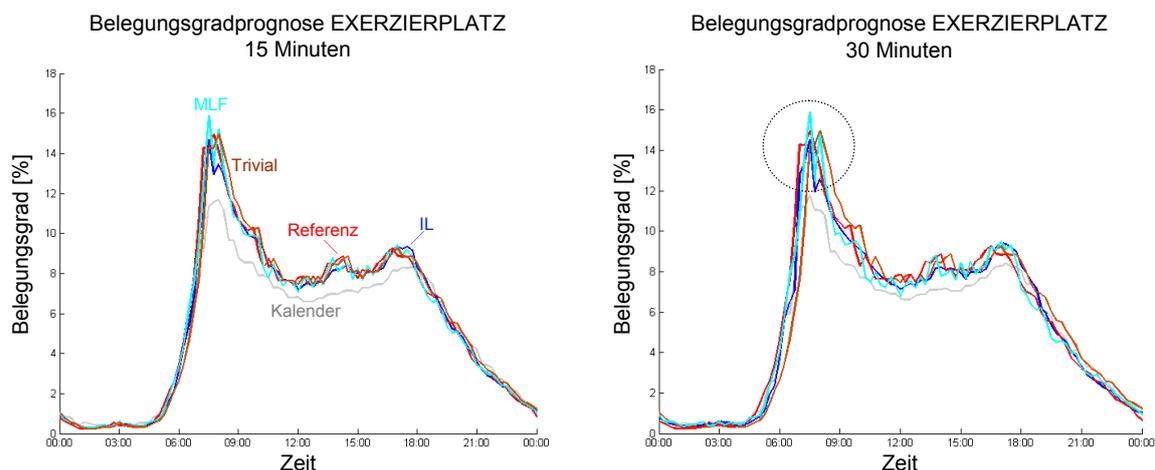


Abb. 5.12 Darstellung von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 30 Minuten (rechts), EXERZIERPLATZ

Für einen Prognosehorizont von 15 Minuten liefern die Instanzbasierten Lernverfahren und die MLF vergleichbar gute Ergebnisse. Die Performanz ist auch bei diesem kurzen Prognosehorizont leicht besser als die der Trivialprognose, was durch das instationäre Verhalten der Ganglinie erklärt werden kann. Die bessere Performanz gilt sowohl für die Prognose mit Angleich an die aktuelle Ganglinie, als auch für die rein mustererkennungsbasierte Prognose. Die Performanz ist erwartungsgemäß deutlich besser als die der kalenderbasierten Prognose.

Das Instanzbasierte Lernverfahren ohne Berücksichtigung der Daten des Detektors EXERZIERPLATZ selber (Szenario "Messwertausfall") liefert bis zu einem Prognosehorizont von 45 Minuten nur leicht bessere Ergebnisse als die kalenderbasierte Prognose. Das MLF

ohne Berücksichtigung der Daten des Detektors EXERZIERPLATZ liefert etwas schlechtere Ergebnisse als die kalenderbasierte Prognose.

Für einen Prognosehorizont von 30 und 45 Minuten liegt die Performanz des Instanzbasierten Lernens leicht über der der beiden MLF, insgesamt sind die Ergebnisse jedoch vergleichbar und jeweils besser als die der Trivialprognose und der kalenderbasierten Prognose.

Bei einem Prognosehorizont von 60 Minuten ist die kalenderbasierte Prognose (und sind somit die Instanzbasierten Lernverfahren entsprechend der verwendeten Einstellung) der Trivialprognose deutlich überlegen. Den geringsten Prognosefehler liefern die MLF. Dies deutet darauf hin, dass die Berücksichtigung der aktuellen Verkehrsdaten auch für einen Horizont von einer Stunde einen Mehrwert bringen kann.

In Abb. 5.13 sind die Zusammenhänge zwischen gemessenen und den prognostizierten Belegungsgraden für Prognosehorizonte von 15 bzw. 45 Minuten grafisch dargestellt. Links ist zu erkennen, dass die Trivialprognose, das Instanzbasierte Lernen ohne Berücksichtigung Daten des Detektors EXERZIERPLATZ und die kalenderbasierte Prognose im Vergleich zu den anderen Verfahren ungünstigere Ergebnisse liefern. Rechts ist zu sehen, dass sich die Trivialprognose (braun) für die 45-Minuten-Prognose deutlich ungünstiger verhält als die anderen Verfahren.

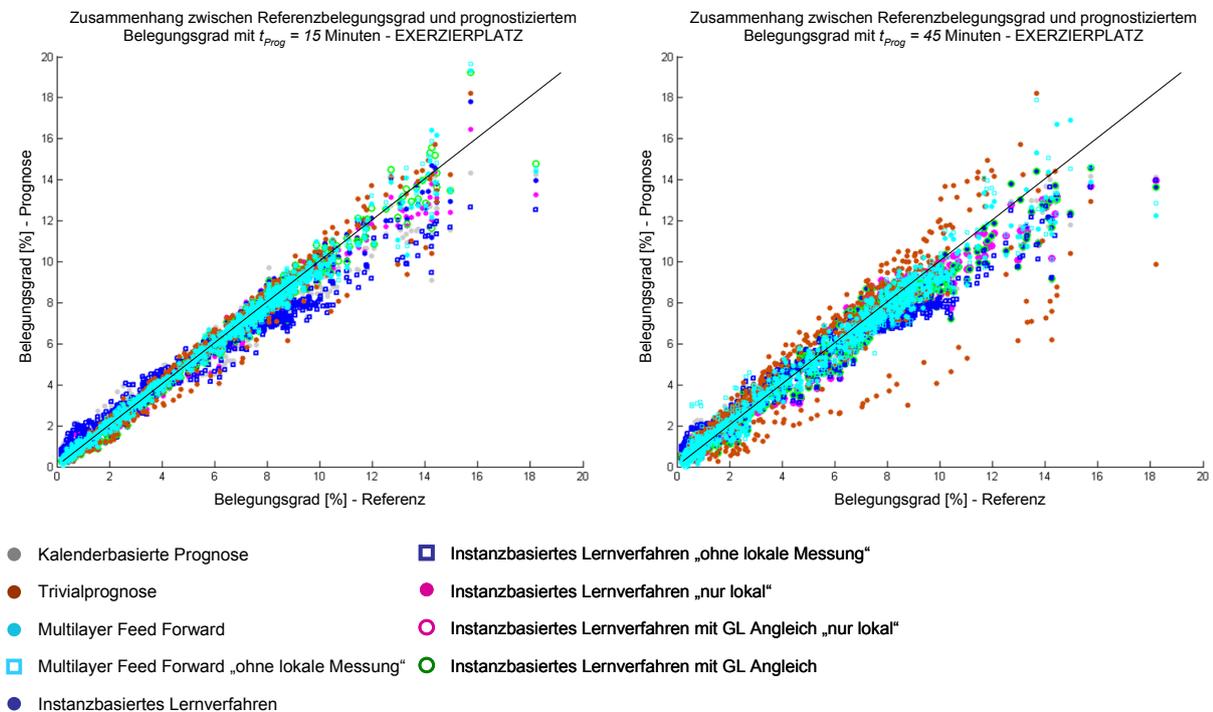


Abb. 5.13 Vergleich von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 45 Minuten (rechts), EXERZIERPLATZ

Abb. 5.14 zeigt den mittleren proportionalen Fehler (links) und den mittleren quadratischen Fehler (rechts) als Netzgrafik. Deutlich zu erkennen ist die schlechter werdende Performanz der Trivialprognose für größer werdende Prognosehorizonte.

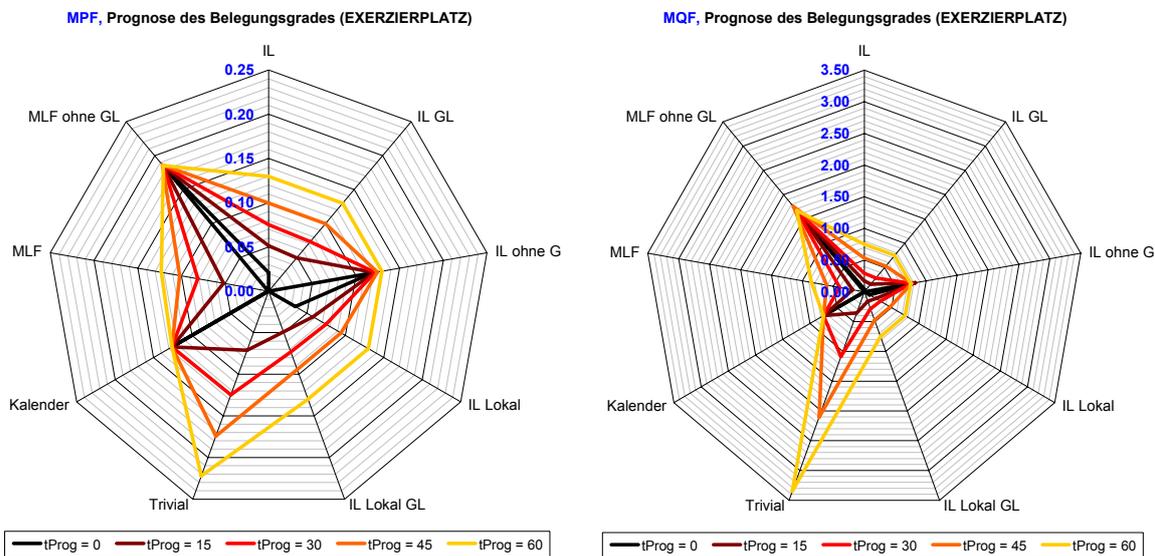


Abb. 5.14 MPF (links) und MQF (rechts) zwischen Prognose und Referenzwert, EXERZIERPLATZ

Tab. 5.5 zeigt die Korrelationsmaße zwischen prognostiziertem Wert und Referenz für die verschiedenen Prognosehorizonte, in Tab. 5.6 sind die mittleren absoluten Fehler der Prognose dargestellt.

Verfahren	t _{Prog} Gütemaß	r ₂				
		0	15	30	45	60
Instanzbasiertes Lernen IL	-	-	0.994	0.990	0.985	0.985
IL mit GL Anpassung	-	-	0.994	0.990	0.985	0.985
IL ohne GL	0.972	0.972	0.972	0.976	0.977	0.986
IL Lokal	-	-	0.993	0.990	0.986	0.985
IL Lokal mit GL Anpassung	-	-	0.994	0.990	0.986	0.985
Trivialprognose	-	-	0.988	0.962	0.926	0.882
Kalenderbasierte Prognose	0.985	0.985	0.985	0.985	0.985	0.985
MLF 1	-	-	0.994	0.989	0.980	0.971
MLF 2	-	-	0.993	0.986	0.979	0.971
MLF 1 ohne GL	0.958	0.958	0.953	0.945	0.939	0.945

Tab. 5.5 Korrelation zwischen Prognose und Referenz, Belegungsgrad, EXERZIERPLATZ

Verfahren	t_{Prog} Fehlermaß	0 MAF	15 MAF	30 MAF	45 MAF	60 MAF
Instanzbasiertes Lernen IL	-	-	0.26	0.38	0.51	0.66
IL mit GL Anpassung	-	-	0.25	0.37	0.51	0.66
IL ohne GL	-	0.60	0.61	0.62	0.63	0.66
IL Lokal	-	-	0.29	0.38	0.48	0.66
IL Lokal mit GL Anpassung	-	-	0.25	0.37	0.48	0.66
Trivialprognose	-	-	0.36	0.63	0.88	1.13
Kalenderbasierte Prognose	-	0.66	-	-	-	-
MLF 1	-	-	0.26	0.41	0.52	0.62
MLF 2	-	-	0.25	0.49	0.51	0.60
MLF 1 ohne GL	-	0.93	0.93	0.94	0.94	0.94

Tab. 5.6 MAF zwischen Prognosewert und Referenz, Belegungsgrad, EXERZIERPLATZ

ERZHERZOG (1.-7. Dezember 2007)

Die Belegungsganglinie am Detektor ERZHERZOG zeigt üblicherweise weniger stark ausgeprägte Spitzen als die Ganglinie am EXERZIERPLATZ (siehe Abb. 5.15 für ein Beispiel).

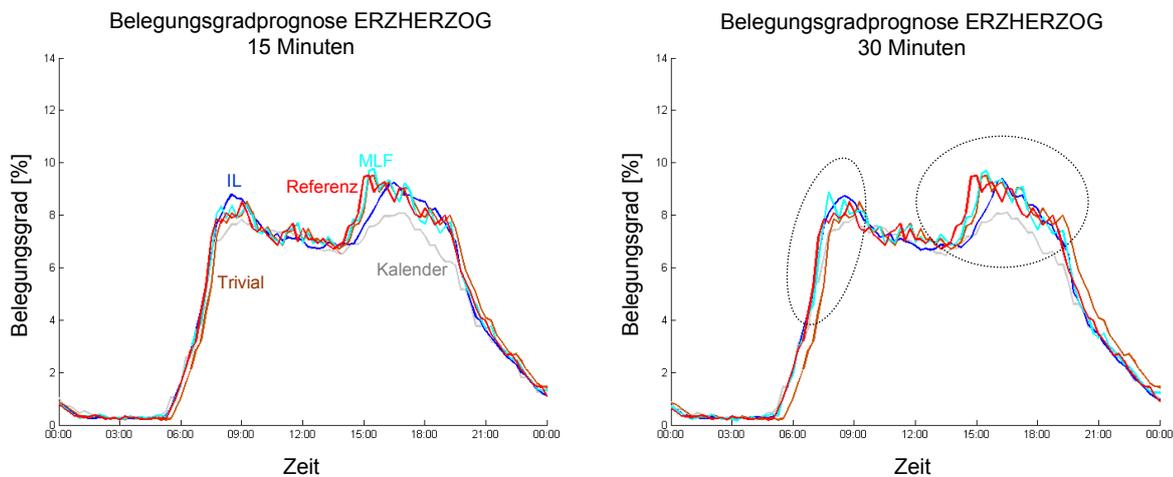


Abb. 5.15 Darstellung von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 30 Minuten (rechts), ERZHERZOG

Für einen Prognosehorizont von 15 Minuten liefern die Instanzbasierten Lernverfahren mit Ganglinien-Angleich und die MLF die beste Performanz. Die Trivialprognose ist für diesen Fall des kurzen Prognosehorizonts besser als die Instanzbasierten Lernverfahren ohne Angleich der Prognose an den aktuellen Messwert.

Instanzbasierte Lernverfahren und MLF ohne Berücksichtigung der Zeitreihe am betrachteten Detektor ERZHERZOG als unabhängige Variable liefern vergleichbare Ergebnisse (außer

für einen Prognosehorizont von 60 Minuten, bei dem offensichtlich wieder die Berücksichtigung der aktuellen Messwerte im MLF den Ausschlag gibt).

Die Trivialprognose ist ab einem Prognosehorizont von 30 Minuten den mustererkennungsbasierten Verfahren und ab einem Horizont von 45 Minuten auch dem kalenderbasierten Verfahren unterlegen. Sämtliche Verfahren unter Berücksichtigung aktueller Messwerte verhalten sich hier günstiger als die kalenderbasierte Prognose.

In Abb. 5.16 sind die Zusammenhänge zwischen gemessenen und den prognostizierten Belegungsgraden für Prognosehorizonte von 15 bzw. 45 Minuten grafisch dargestellt. Links (15 Minuten) ist zu erkennen, dass die kalenderbasierte Prognose die größten Abweichungen aufweist, rechts (45 Minuten) ist zu sehen, dass die Trivialprognose deutlich ungünstigere Ergebnisse liefert als die anderen Verfahren.

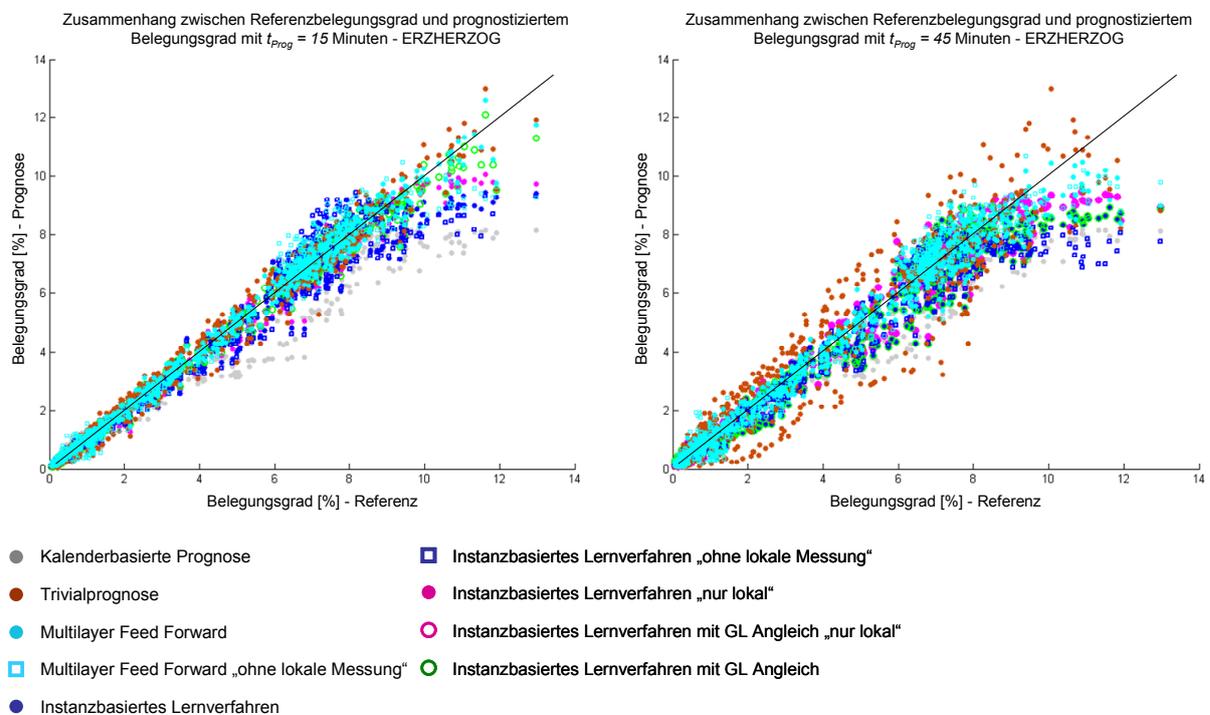


Abb. 5.16 Vergleich von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 45 Minuten (rechts), ERZHERZOG

Abb. 5.17 zeigt den mittleren proportionalen Fehler (links) und den mittleren quadratischen Fehler (rechts) als Netzgrafik. Zu erkennen ist in beiden, dass die MLF insgesamt die beste Performanz liefern. Tab. 5.7 zeigt die Korrelationsmaße zwischen prognostiziertem Wert und Referenz für die verschiedenen Prognosehorizonte, in Tab. 5.8 sind die mittleren absoluten Fehler der Prognose dargestellt.

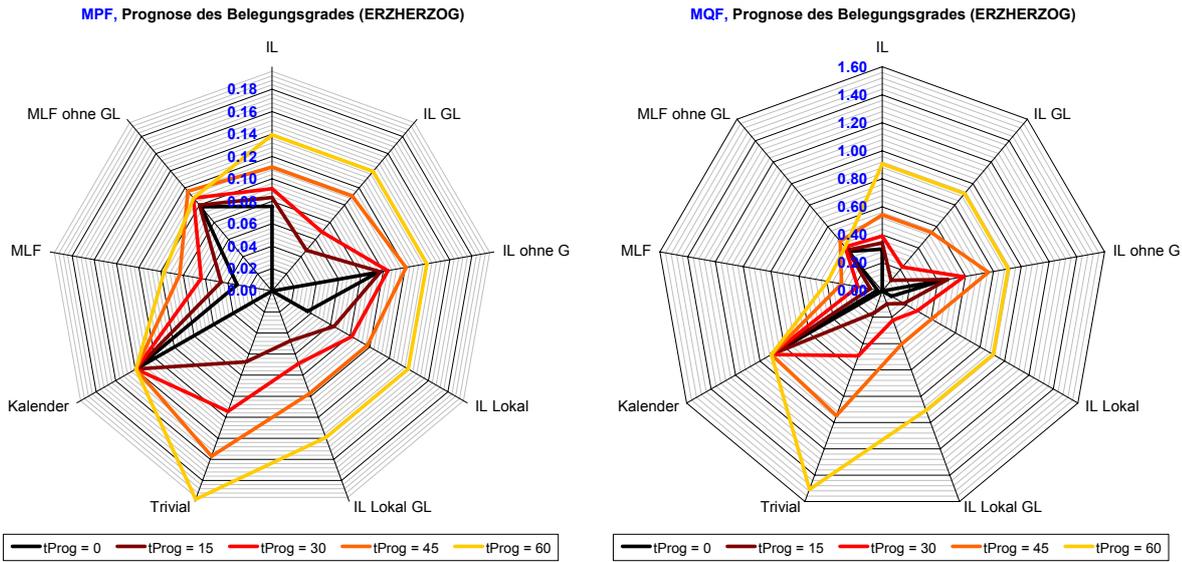


Abb. 5.17 MPF (li.) und MQF (re.) zwischen Prognose und Referenz, ERZHERZOG

Verfahren	t _{Prog} Gütemaß	0	15	30	45	60
		r ₂				
Instanzbasiertes Lernen IL	-	-	0.985	0.983	0.978	0.970
IL mit GL Anpassung	-	-	0.995	0.990	0.978	0.970
IL ohne GL	0.980	0.979	0.974	0.970	0.970	
IL Lokal	-	0.992	0.987	0.983	0.970	
IL Lokal mit GL Anpassung	-	0.995	0.990	0.983	0.970	
Trivialprognose	-	0.992	0.976	0.954	0.926	
Kalenderbasierte Prognose	0.970	-	-	-	-	
MLF 1	-	0.996	0.991	0.986	0.981	
MLF 2	-	0.996	0.991	0.982	0.983	
MLF 1 ohne GL	0.985	0.985	0.984	0.982	0.984	

Tab. 5.7 Korrelation zwischen Prognosewert und Referenz, Belegungsgrad, ERZHERZOG

Verfahren	t _{Prog} Fehlermaß	0	15	30	45	60
		MAF	MAF	MAF	MAF	MAF
Instanzbasiertes Lernen IL	-	-	0.36	0.40	0.48	0.61
IL mit GL Anpassung	-	-	0.21	0.30	0.48	0.61
IL ohne GL	0.41	0.42	0.45	0.52	0.61	
IL Lokal	-	0.28	0.35	0.42	0.61	
IL Lokal mit GL Anpassung	-	0.21	0.30	0.42	0.61	
Trivialprognose	-	0.29	0.47	0.68	0.86	
Kalenderbasierte Prognose	0.61	-	-	-	-	
MLF 1	-	0.20	0.28	0.36	0.42	
MLF 2	-	0.20	0.28	0.40	0.39	
MLF 1 ohne GL	0.40	0.42	0.46	0.50	0.46	

Tab. 5.8 MAF zwischen Prognosewert und Referenz, Belegungsgrad, ERZHERZOG

Abhängigkeit vom Umfang der Trainingsdatenbasis

Die im vorherigen Abschnitt vorgestellten Ergebnisse sind auf Basis des "vollständigen" Datenarchivs ermittelt worden. Vollständig bedeutet hier, dass alle zur Verfügung stehenden Beobachtungen genutzt wurden und nicht, dass für jedes Zeitintervall eine Beobachtung zur Verfügung stand.

Es ist zu erwarten, dass die Prognosegüte bei abnehmendem Umfang der Trainingsdatenbasis sinkt. Theoretisch kann in einer Trainingsdatenbasis jede denkbare Verkehrssituation vorliegen, was die Prognose in untypischen Situationen erleichtern würde. Praktisch ist die Anzahl der Instanzen beschränkt auf die Archivierungsdauer und durch die Beobachtungshäufigkeit der abhängigen Variablen.

Der Einfluss des Umfangs der Trainingsdatenbasis wird am Beispiel der Schätzung lokaler Belegungsgrade des Detektors "ERZHERZOG" untersucht (siehe Abb. 5.18). Erwartungsgemäß reduzieren sich die Fehlermaße und steigt die Korrelation der Schätzung und der Referenz mit wachsender Anzahl der zur Schätzung verfügbaren Instanzen.

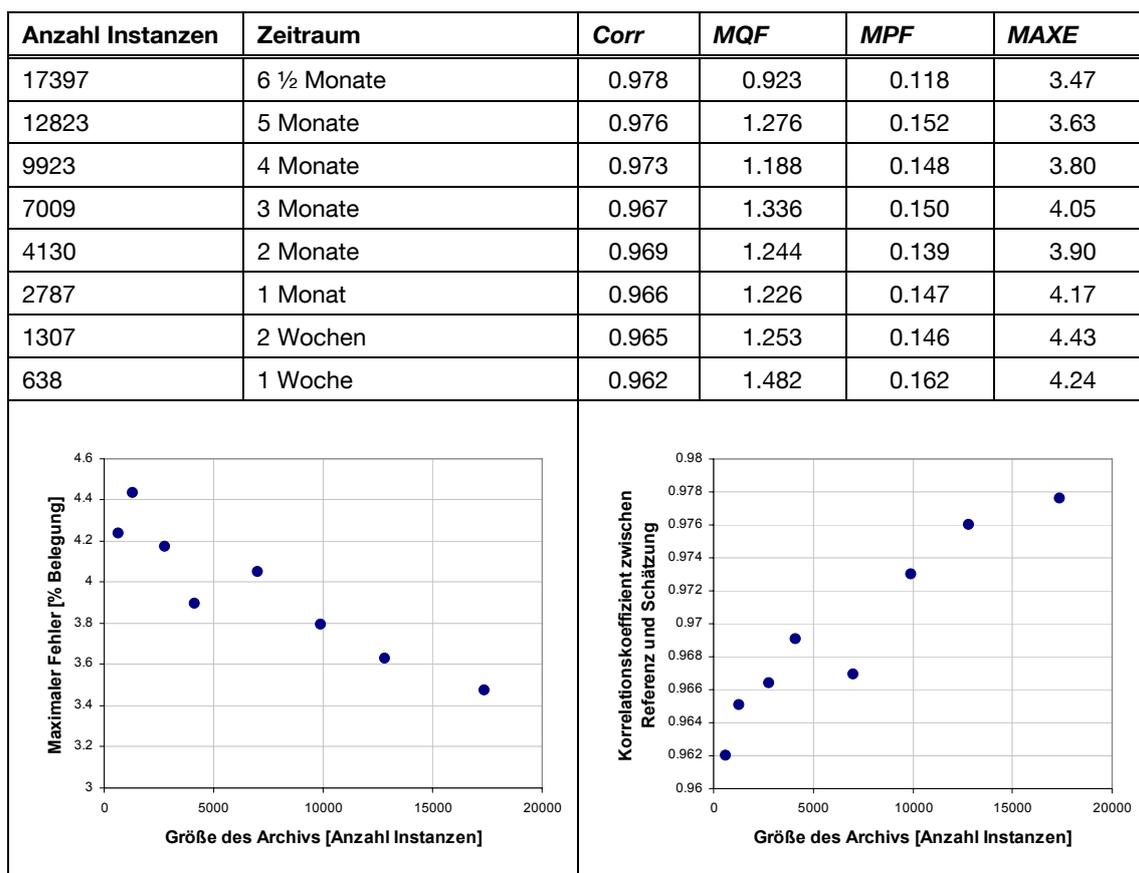


Abb. 5.18 Einfluss der Trainingsdatenbasis auf das Schätzergebnis

Für verschiedene Anwendungen wäre es interessant zu wissen, wieviele Beobachtungen für eine ausreichend zutreffende Prognose benötigt werden. Denkbar ist beispielsweise die temporäre Erhebung von lokalen Verkehrsdaten mit mobilen Detektoren. Entsprechende Systeme übermitteln die erfassten Daten drahtlos (zum Beispiel über GPRS) an eine Zentrale und sind somit unabhängig von Kommunikationsinfrastrukturen, so dass sie mit vergleichsweise geringem Aufwand installiert werden können. Ein derartiges System könnte an verschiedenen neuralgischen Punkten im Straßennetz zum Aufbau einer Instanzbasis installiert werden. Gleiches gilt für die Durchführung einer Messkampagne mittels mobil einsetzbarer Kennzeichenerfassungssysteme. Möglicherweise genügen schon einige wenige Tage zur Erfassung eines relativ vollständigen Sets, so dass anhand der kontinuierlich erfassten Daten umliegender Detektoren eine durchgängige Schätzung ermöglicht wird. Natürlich sind diese Ergebnisse lokationsspezifisch und daher exemplarisch zu sehen und hängen stark davon ab, inwieweit sich die Muster täglich bzw. wöchentlich wiederholen.

5.6 Ergebnisse der Reisezeitprognose

In diesem Abschnitt werden die Ergebnisse der Reisezeitprognose beschrieben. Die mittlere absolute Abweichung mit Toleranz und der Korrelationskoeffizient als Übereinstimmungsmaß sind tabellarisch aufgelistet, daneben sind Zeitreihen und Scatter-Plots zur Visualisierung der Ergebnisse dargestellt. In Anhang 3 finden sich weitere Darstellungen.

OPERNRING (Basis: Taxireisezeiten, 16.-22. November 2007)

Auf dem OPERNRING ist die Abdeckung mit Taxireisezeiten relativ hoch (siehe Abb. 5.19 in rot). In blau (IL) und cyan (MLF) dargestellt sind die Ergebnisse der Reisezeitprognose auf Basis der Belegungsgrade von Detektoren auf und stromaufwärts der Route OPERNRING für Horizonte von 0 Minuten (oben) und 15 Minuten (unten).

Es sind die Reisezeiten über eine Woche (16.-22. November 2007) dargestellt. Am ersten Tag (Situation a) sind extrem hohe Taxireisezeit zu beobachten. Da diese von mehreren aufeinander folgenden Taxis "bestätigt" werden, ist davon auszugehen, dass hier tatsächlich eine substantielle Störung im Verkehrsablauf vorgelegen hat. Es ist zu sehen, dass sowohl das Instanzbasierte Lernverfahren als auch das MLF den deutlichen Reisezeitanstieg abbilden, jedoch nicht in der Ausprägung wie durch die Taxis beobachtet. Im weiteren Verlauf zeigt sich – bis auf einen Ausreißer des Instanzbasierten Lernverfahrens in Situation b – ein plausibler Verlauf der Reisezeit. Insbesondere konnten von beiden Verfahren die ausgeprägten Reisezeitanstiege am Montag und Dienstag (Situationen c und d) im Vergleich zu Mittwoch und Donnerstag abbilden (für einen Prognosehorizont von 0 Minute erwartungsgemäß besser als für einen Horizont von 15 Minuten).

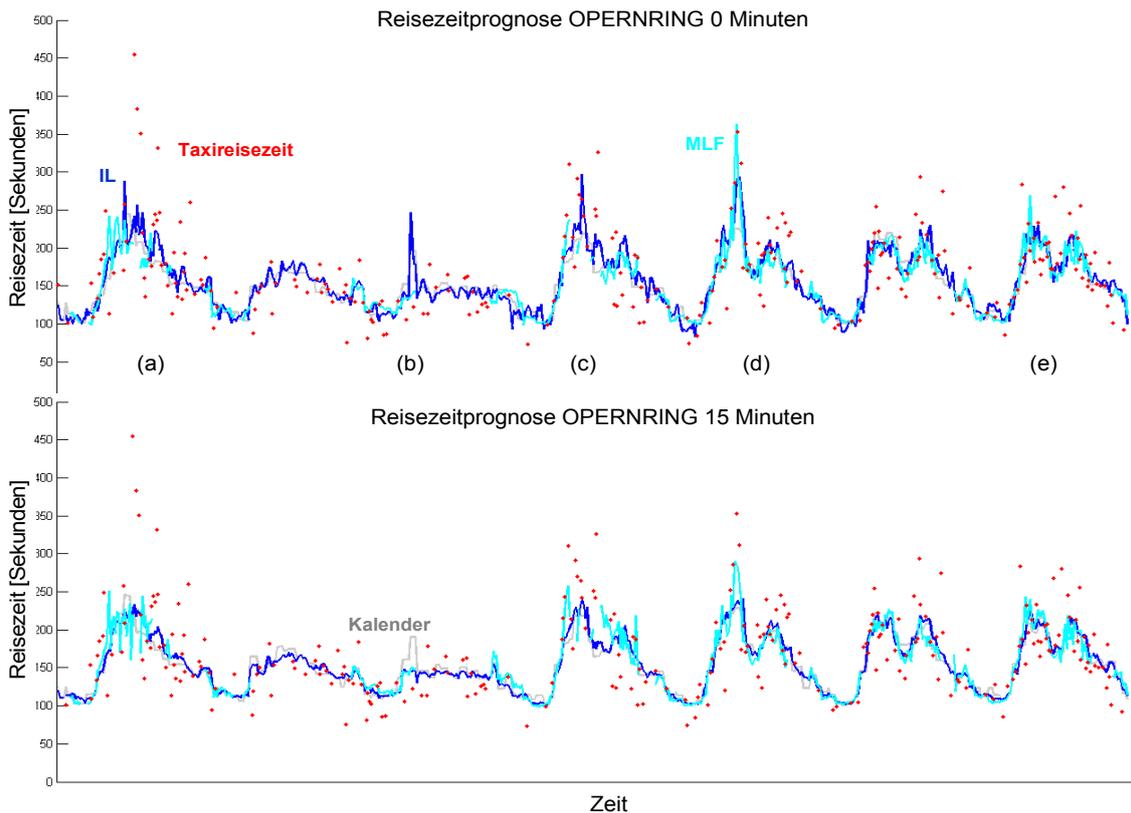


Abb. 5.19 Verlauf der prognostizierten Reisezeiten und zugehörige Taxireisezeiten für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), OPERNRING

Abb. 5.20 zeigt den Zusammenhang zwischen prognostizierter Reisezeit und Taxireisezeit für einen Horizont von 0 Minuten (links) bzw. 30 Minuten (rechts). Deutlich zu erkennen ist, dass die geschätzte Reisezeit (links) insbesondere für hohe Reisezeiten deutlich näher an der Taxireisezeit liegt als für die 30-Minuten-Prognose (rechts). Im Falle der Schätzung konnten plausible Abweichungen der unter Berücksichtigung der aktuellen Verkehrsdaten ermittelten Reisezeiten von der kalenderbasierten Schätzung beobachtet werden. Für die 30 Minuten-Prognose liegen die Prognose auf Basis der aktuellen Messwerte und die kalenderbasierte Schätzung nahe zusammen.

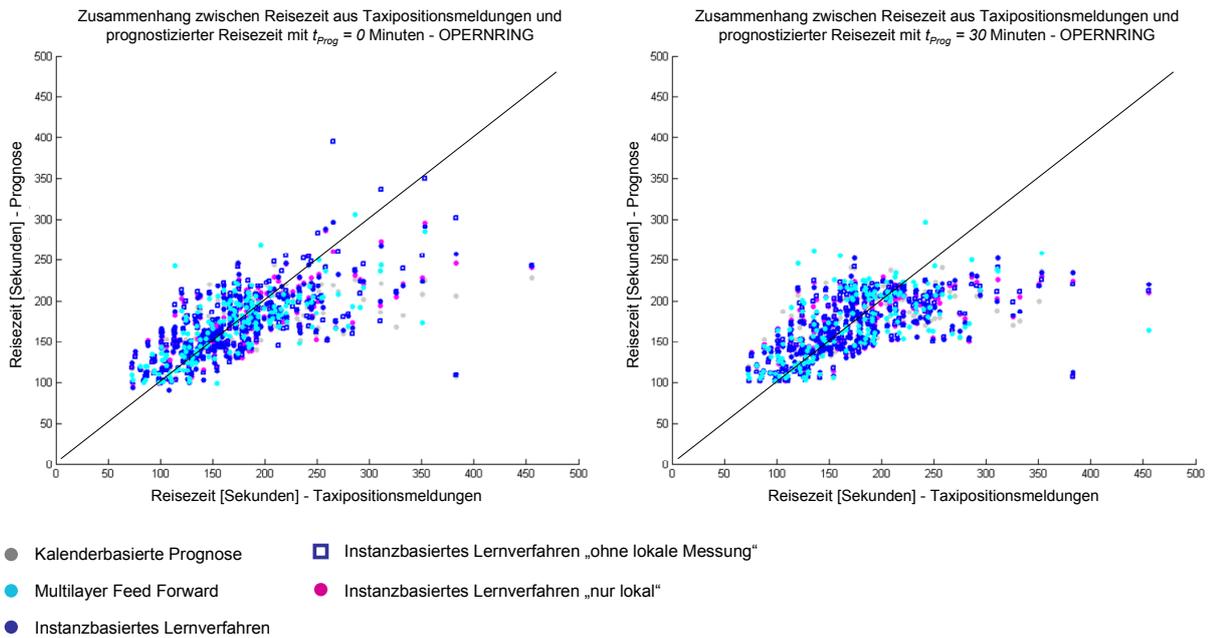


Abb. 5.20 Vergleich von prognostizierter Reisezeit und Taxireisezeiten für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), OPERNRING

In Tab. 5.9 sind die Korrelationen zwischen prognostizierten Reisezeiten und Taxireisezeiten dargestellt. Tab. 5.10 zeigt die mittleren absoluten Abweichungen mit einer zugelassenen Toleranz von 60 Sekunden zwischen prognostizierten Reisezeiten und Taxireisezeiten.

Insgesamt sind die Übereinstimmungen zwischen prognostizierten Reisezeiten und Taxireisezeiten für die Instanzbasierten Lernverfahren (Vorteile insbesondere für 30 und 45 Minuten Horizont) und die MLF (Vorteile für 0 und 15 Minuten Horizont) jeweils größer als für die rein kalenderbasierte Prognose.

Verfahren	Übereinstimmungsmaß	t_{Prog}				
		0 r_2	15 r_2	30 r_2	45 r_2	60 r_2
Instanzbasiertes Lernen IL		0.68	0.64	0.63	0.63	0.59
IL ohne GL		0.67	0.65	0.62	0.62	0.59
IL Lokal		0.68	0.65	0.62	0.62	0.59
Kalenderbasierte Prognose		0.59				
MLF 1		0.72	0.71	0.60	0.60	0.58

Tab. 5.9 Korrelation zwischen Prognosewert und Taxireisezeit, OPERNRING

Verfahren	Abweichungsmaß	t_{prog}				
		0	15	30	45	60
		MAA _T [s]				
Instanzbasiertes Lernen IL		4	5	5	5	6
IL ohne GL		4	5	5	5	6
IL Lokal		4	5	5	5	6
Kalenderbasierte Prognose		6				
MLF 1		3	3	5	5	6

Tab. 5.10 Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Taxireisezeit, OPERNRING (tolerierte Abweichung: 60 Sekunden)

KEPPLER-WEST (Basis: Taxireisezeiten, 16.-22. November 2007)

Auf der Route KEPPLER-WEST sind deutlich weniger Taxireisezeiten zu beobachten als auf der Route OPERNRING (siehe Abb.5.21).

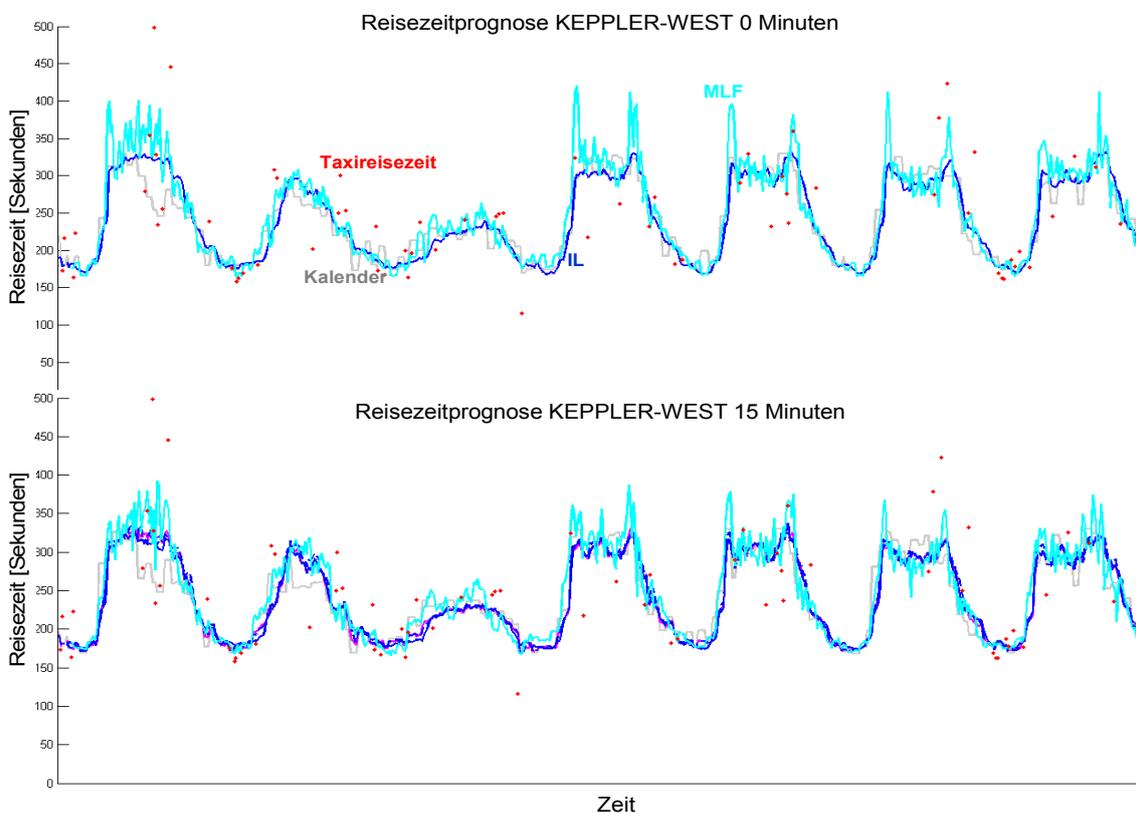


Abb.5.21 Verlauf der prognostizierten Reisezeiten und zugehörige Taxireisezeiten für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), OPERNRING

Insgesamt erscheinen die prognostizierten Reisezeiten plausibel. Abb. 5.22, sowie Tab. 5.11 und Tab. 5.12 zeigen, dass die mittels Instanzbasiertem Lernen und MLF prognostizierten Reisezeiten besser mit den Taxireisezeiten übereinstimmen als die rein kalenderbasierten Reisezeitschätzungen.

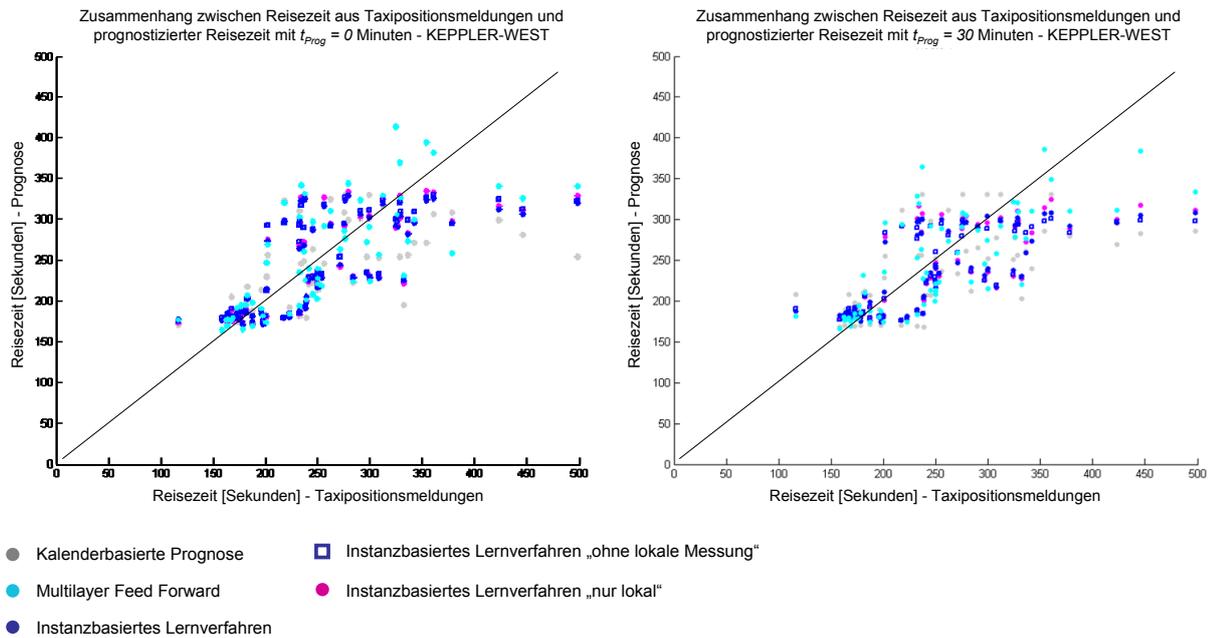


Abb. 5.22 Vergleich von prognostizierter Reisezeit und Taxireisezeiten für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), KEPLER-WEST

Verfahren	Übereinstimmungsmaß	t_{Prog}				
		0	15	30	45	60
		r_2	r_2	r_2	r_2	r_2
Instanzbasiertes Lernen IL		0.74	0.73	0.73	0.72	0.62
IL ohne GL		0.75	0.74	0.72	0.73	0.62
IL Lokal		0.75	0.73	0.73	0.71	0.62
Kalenderbasierte Prognose		0.62				
MLF 1		0.74	0.73	0.75	0.78	0.75

Tab. 5.11 Korrelation zwischen Prognosewert und Taxireisezeit, KEPLER-WEST"

Verfahren	Abweichungsmaß	t_{Prog}				
		0	15	30	45	60
		MAA_T [s]				
Instanzbasiertes Lernen IL		6	7	7	7	10
IL ohne GL		6	7	7	8	10
IL Lokal		6	7	7	8	10
Kalenderbasierte Prognose		10				
MLF 1		7	7	6	6	8

Tab. 5.12 Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Taxireisezeit, KEPLER-WEST (tolerierte Abweichung: 60 Sekunden)

INGOLSTÄDTER (Basis: Fahrzeugwiedererkennung, 16.-22. November 2007)

Auf der INGOLSTÄDTER wurden als Trainings- und Referenzdaten Reisezeiten aus Fahrzeugwiedererkennung verwendet. Als unabhängige Variable wurden Verkehrsstärken genutzt. Abb. 5.23 zeigt den Verlauf der prognostizierten Reisezeiten (0 Minuten oben und 15 Minuten unten) für das Instanzbasierte Lernverfahren und ein MLF im Vergleich zu den Reisezeiten aus Fahrzeugwiedererkennung. Insgesamt ist eine gute Übereinstimmung und ein

plausibler Verlauf festzustellen. Insbesondere der deutlich überdurchschnittliche Anstieg der Reisezeit am 22. November (Situation d ganz rechts in Abb. 5.23) wird vom Instanzbasierten Lernverfahren gut abgebildet. Wiederkehrende Reisezeitverläufe werden ebenfalls gut abgebildet (z.B. Situation a und b). Am vorletzten Tag in der Morgenspitze (Situation c) bildet das MLF die Reisezeiten auf Basis von Fahrzeugwiedererkennung etwas besser ab als das Instanzbasierte Lernverfahren.

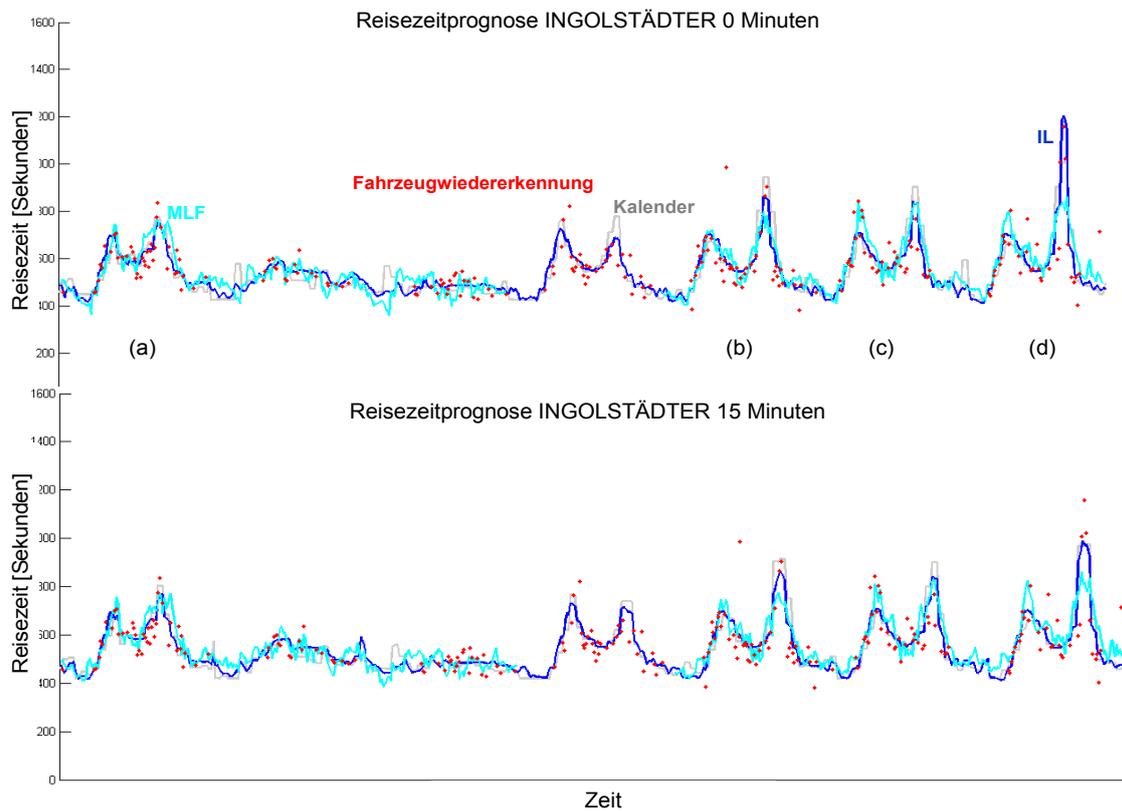


Abb. 5.23 Verlauf der prognostizierten Reisezeiten und zugehörige Reisezeiten aus Fahrzeugwiedererkennung für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), INGOLSTÄDTER

Abb. 5.26 zeigt den ausgeprägten Zusammenhang zwischen 0 Minuten (links) bzw. 30 Minuten (rechts) Prognose und Reisezeiten aus Fahrzeugwiedererkennung.

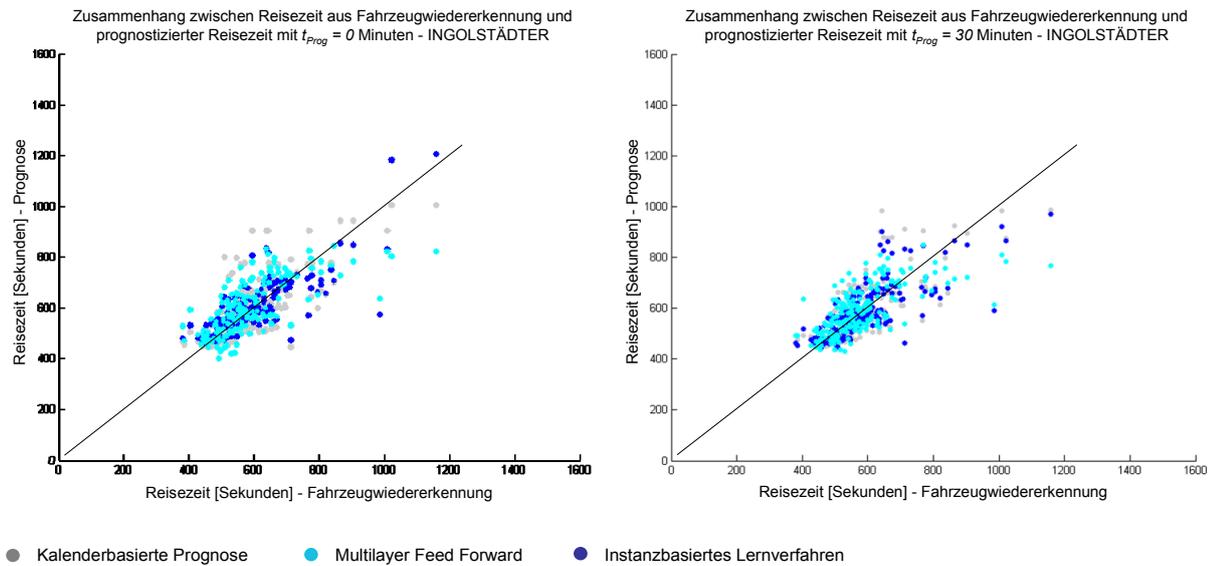


Abb. 5.24 Vergleich von prognostizierter Reisezeit und Reisezeiten aus Fahrzeugwiedererkennung für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), INGOLSTÄDTER

In Tab. 5.13 und Tab. 5.14 sind die Korrelationen bzw. die mittleren absoluten Abweichungen zwischen der prognostizierten Reisezeit und der Reisezeit aus Fahrzeugwiedererkennung mit einer zugelassenen Toleranz von 60 Sekunden dargestellt. Wie zu sehen ist, ist die Übereinstimmung zwischen Instanzbasiertem Lernen und den Reisezeiten aus Fahrzeugwiedererkennung etwas stärker als die zwischen kalenderbasierter Schätzung bzw. MLF und Fahrzeugwiedererkennung.

Verfahren	t_{Prog}				
	0	15	30	45	60
Übereinstimmungsmaß	r_2	r_2	r_2	r_2	r_2
Instanzbasiertes Lernen IL	0.81	0.78	0.77	0.75	0.72
Kalenderbasierte Prognose	0.72				
MLF 1	0.72	0.70	0.67	0.66	0.66

Tab. 5.13 Korrelation zwischen Prognosewert und Reisezeit auf Basis von Fahrzeugwiedererkennung, INGOLSTÄDTER

Verfahren	t_{Prog}				
	0	15	30	45	60
Abweichungsmaß	MAA_T [s]	MAA_T [s]	MAA_T [s]	MAA_T [s]	MAA_T [s]
Instanzbasiertes Lernen IL	11	13	15	16	19
Kalenderbasierte Prognose	19				
MLF 1	21	22	23	23	21

Tab. 5.14 Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Reisezeit auf Basis von Fahrzeugwiedererkennung, INGOLSTÄDTER (tolerierte Abweichung: 60 Sekunden)

Zwischenfazit zur Reisezeitprognose

Insgesamt konnten in allen untersuchten Szenarien plausible Reisezeiten prognostiziert werden. Dabei stellt die Schätzung, also die Prognose für einen Horizont von 0 Minuten,

einen relevanten Anwendungsfall dar, den es zunächst zu meistern gilt. Trotz insgesamt aus Sicht des Autors guter Prognoseergebnisse, sind genaue Schätzungen hoher Reisezeiten eine schwieriges Unterfangen. Das hat den folgenden Grund:

Situationen mit untypisch hohen Reisezeiten treten per Definition selten auf, so dass sie auch in den zur Mustererkennung und zum Training verwendeten Datenbasen im Vergleich zu anderen Situationen relativ selten sind. Somit kann es passieren, dass die niedrigen Reisezeiten die k Nächsten Nachbarn und somit den Erwartungswert dominieren. Diese Problematik alleine könnte durch eine entsprechende Datenvorverarbeitung (etwa einer Segmentierung der Datenbasis) möglicherweise gelöst werden. Ein größeres Problem ist, dass der Zusammenhang zwischen lokalen unabhängigen Verkehrsvariablen und der Reisezeit im Bereich hoher Reisezeiten nicht eindeutig ist. Die (mögliche) Streubreite der Reisezeit ist deutlich höher als die der lokalen Verkehrskenngrößen b und q . Diese sind bei (nur theoretisch erreichbaren) 100% bzw. bei der Kapazität begrenzt. Die Reisezeit hingegen kann theoretisch unendlich groß werden und erreicht in der Praxis nicht selten das fünf- oder sechsfache der idealen Reisezeiten. Mehrere sehr ähnliche Muster lokaler Verkehrsdaten können zu stark unterschiedlicher Reisezeiten korrespondieren, womit eine genaue Schätzung des Erwartungswertes nahezu unmöglich wird. Dies betrifft nicht nur das Instanzbasierte Lernen, sondern naturgemäß auch andere Mustererkennungsansätze.

Das Instanzbasierte Lernen bietet aufgrund seiner Struktur verschiedene Möglichkeiten, die beschriebene Problematik zu berücksichtigen. Im Folgenden werden drei Verfahrensergänzungen vorgestellt und exemplarisch angewendet.

5.7 Ergänzungen zur Reisezeitprognose

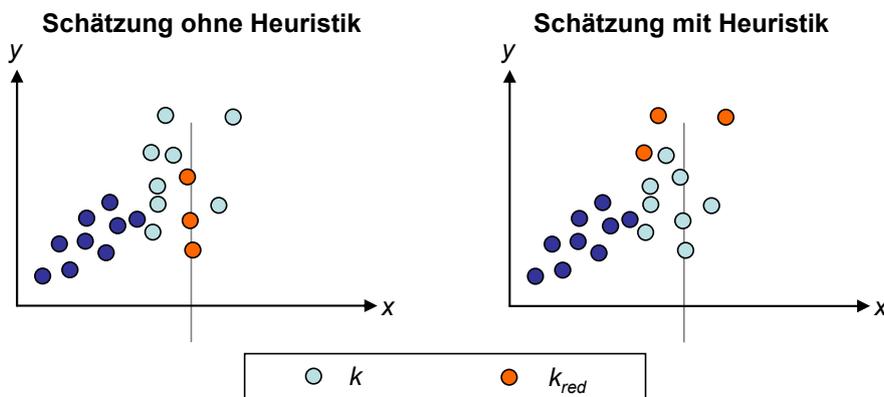
5.7.1 Test von Heuristiken zur Verbesserung der Reisezeitprognose

Da hohe Reisezeiten aufgrund seltener Beobachtung in den k ähnlichsten Instanzen tendenziell unterrepräsentiert sind, werden Heuristiken zur Bevorzugung hoher Reisezeiten innerhalb der k ähnlichsten Instanzen vorgeschlagen und getestet.

Heuristik 1: Bevorzugung größerer Reisezeiten

Befinden sich unter den zehn ähnlichsten historischen Beobachtungen beispielsweise drei im Vergleich zu den anderen relativ hohe Reisezeiten, so könnten diese trotz geringerer Ähnlichkeit höher gewichtet werden. Dieses Vorgehen widerspricht im Prinzip einer grundlegenden Annahme des Instanzbasierten Lernens: "je ähnlicher, desto wahrscheinlicher". Andererseits ist in Bereichen hoher Reisezeiten in der Regel eine der Grundvoraussetzungen verletzt, da die k ähnlichsten Instanzen sich nicht gleichmäßig um das aktuelle Muster verteilen, sondern tendenziell darunter liegen.

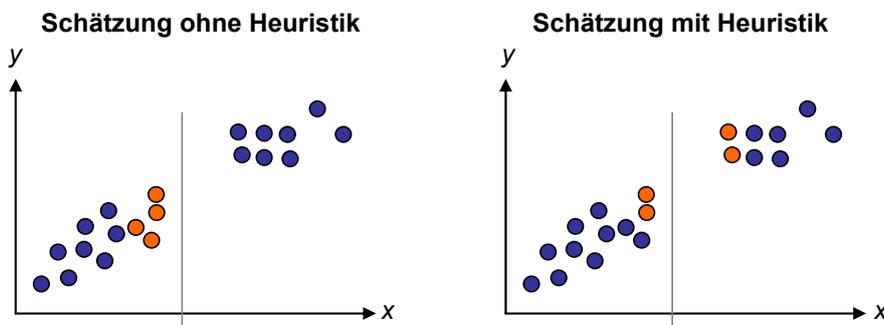
"Nutze die $k_{red,gro\beta}$ größten Reisezeiten der k ähnlichsten Nachbarn zur Ermittlung der Reisezeit, wobei $k_{red,gro\beta} \ll k$." Dabei wird $k = 51$ und $k_{red,gro\beta} = 21$ gewählt.



Heuristik 2: Erzwingen gleichmäßiger Verteilung

Eine andere Möglichkeit ist die Auswahl der k Nächsten Nachbarn entsprechend ihrer Lage um das aktuelle Muster herum. Wenn also beispielsweise die zehn ähnlichsten Muster alle "unter" dem aktuellen Muster liegen, sollten ggf. einige der Instanzen durch darüber liegende Instanzen ersetzt werden, auch wenn diese weniger ähnlich sind.

"Nehme nicht die k ähnlichsten Nachbarn, sondern die $k_{red,vert}/2$ ähnlichsten mit positiver Differenz und die $k_{red,vert}/2$ ähnlichsten mit negativer Differenz." Zum Test wird $k_{red,vert} = 10$ gewählt.



Bei Wahl eines größeren $k_{red,vert}$ besteht die Gefahr, dass die verwendeten Instanzen zu weit entfernt im Merkmalsraum liegen.

Die $k_{red,gro\beta}$ bzw. $k_{red,vert}$ ausgewählten Instanzen werden nicht entsprechend ihrer Distanzen gewichtet, da ansonsten die entfernten Instanzen kaum einen oder keinen Einfluss haben. Stattdessen wird ein einfacher arithmetischer Mittelwert verwendet.

Abb. 5.25 zeigt den Verlauf der geschätzten Reisezeiten ($t_{prog} = 0$ Minuten) für das Instanzbasierte Lernverfahren ohne Heuristik, mit Heuristik 1 und mit Heuristik 2, sowie die Taxireisezeiten. Es ist zu erkennen, dass beide Heuristiken, wie beabsichtigt, tendenziell zu höhe-

ren Reisezeitschätzungen führen als die Schätzung ohne Heuristik. Heuristik 1 scheint die Reisezeit in Schwachverkehrszeiten zu überschätzen. Dies müsste durch entsprechende Regeln unterdrückt werden. Im Falle der deutlich erhöhten Taxireisezeiten um etwa 13:00 Uhr, liegt die Schätzung unter Anwendung von Heuristik 1 näher an den Taxireisezeiten. Heuristik 2 liefert im Bereich niedriger Reisezeiten ähnliche Ergebnisse wie die Schätzung ohne Heuristik. Im Bereich erhöhter Reisezeiten (13:00 Uhr / 16:00 Uhr) werden jedoch höhere (und scheinbar plausiblere) Reisezeiten geschätzt, so dass die Anwendung von Heuristik 2 sinnvoll erscheint.

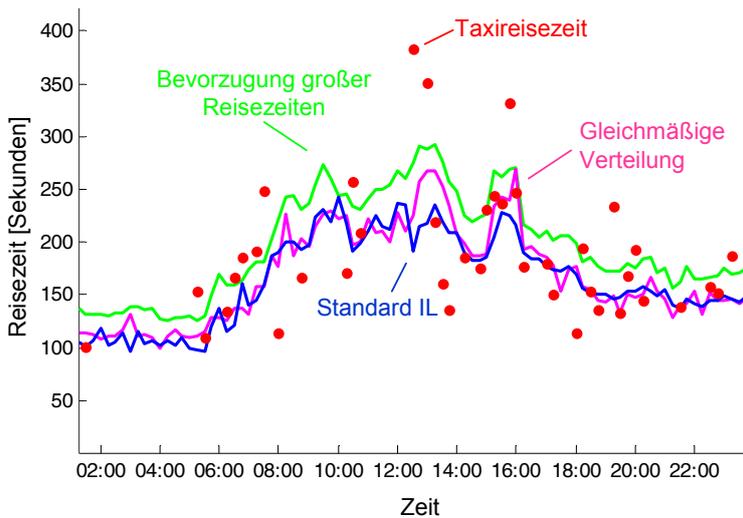


Abb. 5.25 Geschätzte Reisezeit unter Verwendung von Heuristiken (OPERNRING)

5.7.2 Anwendung der Zuverlässigkeitsindikatoren

In Abschnitt 4.5.1 wurden die möglichen Zuverlässigkeitsindikatoren L_x , S_x und σ_y vorgestellt. Diese sollten in unsicheren Schätz- und Prognosesituationen reagieren, was im Folgenden getestet wird.

Abb. 5.26 zeigt die auf Basis des Modells geschätzte Reisezeit und zugehörige Taxireisezeiten. Zusätzlich sind die Indikatoren L_x (unten), S_x (Mitte) und σ_y (oben) dargestellt. Es sei noch einmal darauf hingewiesen, dass die Indikatoren nicht im Nachhinein durch einen Vergleich der Schätzung und der "Referenz" ermittelt werden, sondern online auf Basis der aktuellen Belegungsgrade und den historischen Instanzen geschätzt werden.

Es ist zu erkennen, dass die Indikatoren an den richtigen Stellen reagieren, nämlich dann, wenn die Referenzreisezeit sehr hoch ist und die $k = 51$ nächsten Nachbarn eine niedrigere Reisezeit suggerieren (Situationen (a) bis (e)). L_x und S_x sind in den relevanten Situationen nahezu gespiegelt identisch. Das zeigt, dass in den betreffenden Situationen die Differenz zwischen dem aktuellen Muster und den ähnlichsten historischen Mustern stets oder fast immer negativ ist. Dies wiederum bedeutet, dass die meisten Beobachtungen zu Zeiten we-

niger starker Verkehrsbelastung gemacht wurden, so dass in den Fällen starker Belastung die entsprechenden historischen Beobachtungen unterrepräsentiert sind. Möglicherweise spielt hier eine Rolle, dass Taxifahrer Staus tendenziell eher umfahren und daher zu wenige Referenzbeobachtungen vorliegen. Interessant ist auch die Reaktion von σ_y in Situation (f), in der vom Verfahren vermutlich eine zu hohe Reisezeit geschätzt wird. σ_y beschreibt die Streuung innerhalb der historischen Beobachtungen y . Es ist zu vermuten, dass hier einzelne sehr hohe Reisezeiten den Ausschlag sowohl für das erhöhte σ_y , als auch für die zu hohe Reisezeitschätzung gegeben haben. Das Beispiel demonstriert somit die prinzipielle Eignung der Indikatoren. Für den praktischen Einsatz können die Zuverlässigkeitsindikatoren weiter verfeinert und miteinander zu einer konsistenten Gesamtaussage fusioniert werden.

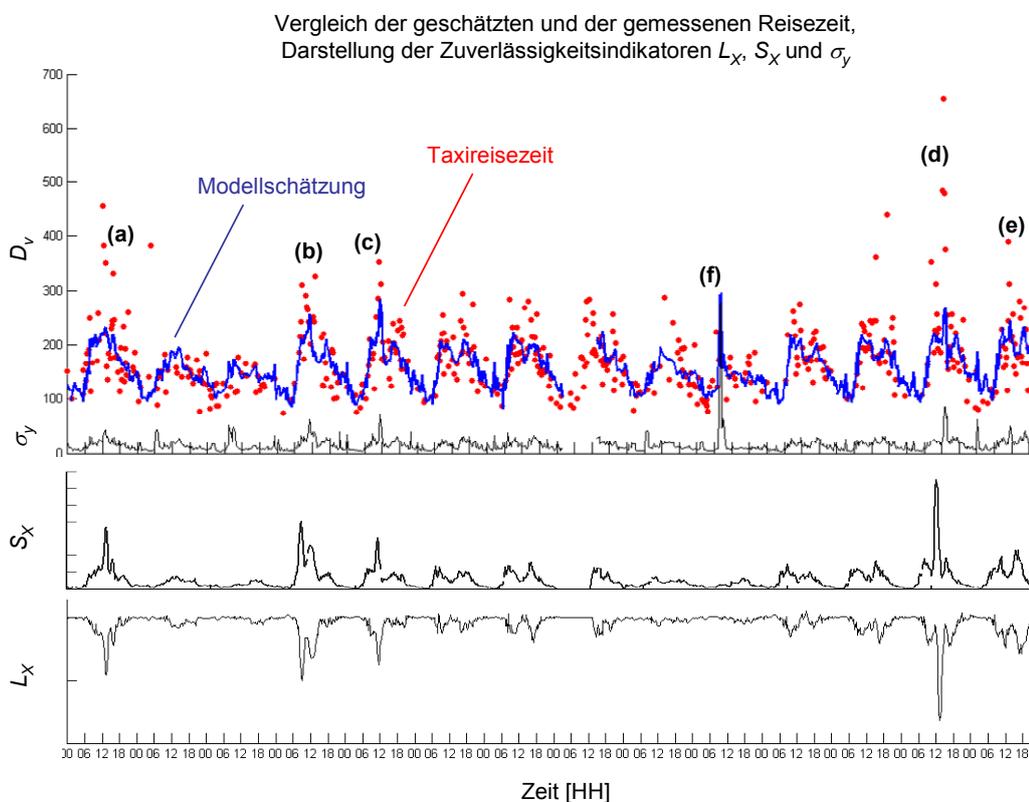


Abb. 5.26 Vergleich von Reisezeitschätzung und Taxireisezeiten ($t_{\text{Prog}} = 0$),
Reaktion der Zuverlässigkeitsindikatoren (OPERNRING)

5.7.3 Abschätzung von Vertrauensintervallen für Reisezeiten

Da es in der Regel nur im Bereich geringer Reisezeiten (und somit weniger relevanten Situationen) möglich ist, eine zuverlässige Schätzung und Prognose abzugeben, wird im Folgenden die in Abschnitt 4.5.2 vorgestellte Methode zur Schätzung von Vertrauensintervallen für Reisezeiten angewandt.

Nützlich erscheint die Abschätzung und Angabe der erwarteten Prognoseunsicherheit in Form von Vertrauensintervallen oder Wahrscheinlichkeitsverteilungen. Anhand dieser kann zwar auch der Erwartungswert ermittelt werden, daneben stehen aber wichtige Kenngrößen zur Verfügung, wie beispielsweise die mit einer bestimmten Wahrscheinlichkeit maximal zu erwartende Reisezeit.

In der vorhandenen Literatur werden zur Prognose von Vertrauensintervallen häufig zeitreihenbasierte Ansätze verwendet. SWIDER ET AL. [2006] prognostizieren Verteilungen von Spotmarktpreisen für Elektrizität unter Einsatz eines ARMA-Modells mit Berücksichtigung externer Variablen X (ARMAX) und eines ARCH (Autoregressive Conditional Heteroscedastic) Ansatzes. Dabei wird die Standardabweichung zeitvariant unter Berücksichtigung aktueller Daten ermittelt. Es wird festgestellt, "*... dass dabei nicht nur der Erwartungswert, sondern auch die Preisverteilung einen erheblichen Einfluss auf die Ergebnisse moderner Methoden zur Entscheidungsunterstützung hat.*" Das gilt ebenso für die Prognose von Verkehrskenngrößen. Entscheidungsunterstützende Anwendungen, wie beispielsweise die zuverlässige Navigation, benötigt Eingangswerte über den Erwartungswert hinaus. TSEKERIS UND STATHOPOULOS [2006] verwenden ein generalisiertes ARCH (GARCH) Modell zur Prognose der Volatilität von lokalen Geschwindigkeiten auf Basis von Detektordaten für 3, 15 und 30 Minuten. KATES ET AL. [2005] generieren die Prognose von Vertrauensintervallen von Geschwindigkeiten durch die Fusion verschiedener Einzelprognosen auf Basis historischer Messwerte und wiederholter stochastischer Realisierungen der Zeitreihenprognose auf Basis eines ARMA(1,1)-Modells.

An dieser Stelle wird das in Abschnitt 4.5.2 dargestellte Konzept zur Ermittlung der Wahrscheinlichkeitsdichten verwendet. Dabei kommt vorteilhaft zum Tragen, dass beim Instanzbasierten Lernen zur Laufzeit sämtliche historischen Beobachtungen zur Verfügung stehen und zur Ermittlung einer Statistik genutzt werden können.

Die erwartete Wahrscheinlichkeitsdichtefunktion der Reisezeit wird aus empirischen Beobachtungen, also Realisierungen der Reisezeit, konstruiert. Im Prinzip ist dies eine Weiterführung des in der Offline Analyse zur Verkehrsqualität üblichen Vorgehens zur Ermittlung von Überlastungswahrscheinlichkeiten (siehe beispielsweise SPANGLER [2008]) aus rein historischen Daten. Es wird jedoch nicht die gesamte, nach Kalenderkriterien klassifizierte historische Stichprobe verwendet, sondern eine Stichprobe auf Basis des aktuellen Verkehrsmusters und Kalendermerkmalen ausgewählt.

Abb. 5.27 zeigt farbig kodierte Überschreitungswahrscheinlichkeiten (Prognosehorizont: 0 Minuten = Schätzung) für die Reisezeit auf dem OPERNRING. Links unten dargestellt sind die auf Basis der Kalendermerkmale wahrscheinlichsten Reisezeiten. Grün dargestellt ist der Erwartungswert (Überschreitungswahrscheinlichkeit = 0.5), blaue Punkte markieren Taxireisezeiten, also die "Referenz". Wie zu sehen ist, liegen im Falle der kalenderbasierten Schätzung in beiden auftretenden kritischen Situationen (a) und (b) die Taxireisezeiten außerhalb

der 10% Kontur. Auf Basis der 51 ähnlichsten Instanzen (verkehrlich und kalenderbasiert) liegen die Taxireisezeiten bis auf einen Fall für Prognosehorizonte von 0 und 15 Minuten (oben) stets unterhalb der 10% Kontur. Die Zutreffenswahrscheinlichkeit nimmt erwartungsgemäß mit wachsendem Horizont ab (30 und 45 Minuten, Mitte), liegt aber noch über der kalenderbasierten Schätzung. Dies zeigt, dass durch die Hinzunahme aktueller verkehrlicher Information ein Mehrwert durch die Schätzung der Überschreitungswahrscheinlichkeiten für verschiedene Prognosehorizonte geliefert werden kann. Der Nutzer wäre in den Fällen (a) und (b) durch die Hinzunahme aktueller Information besser informiert.

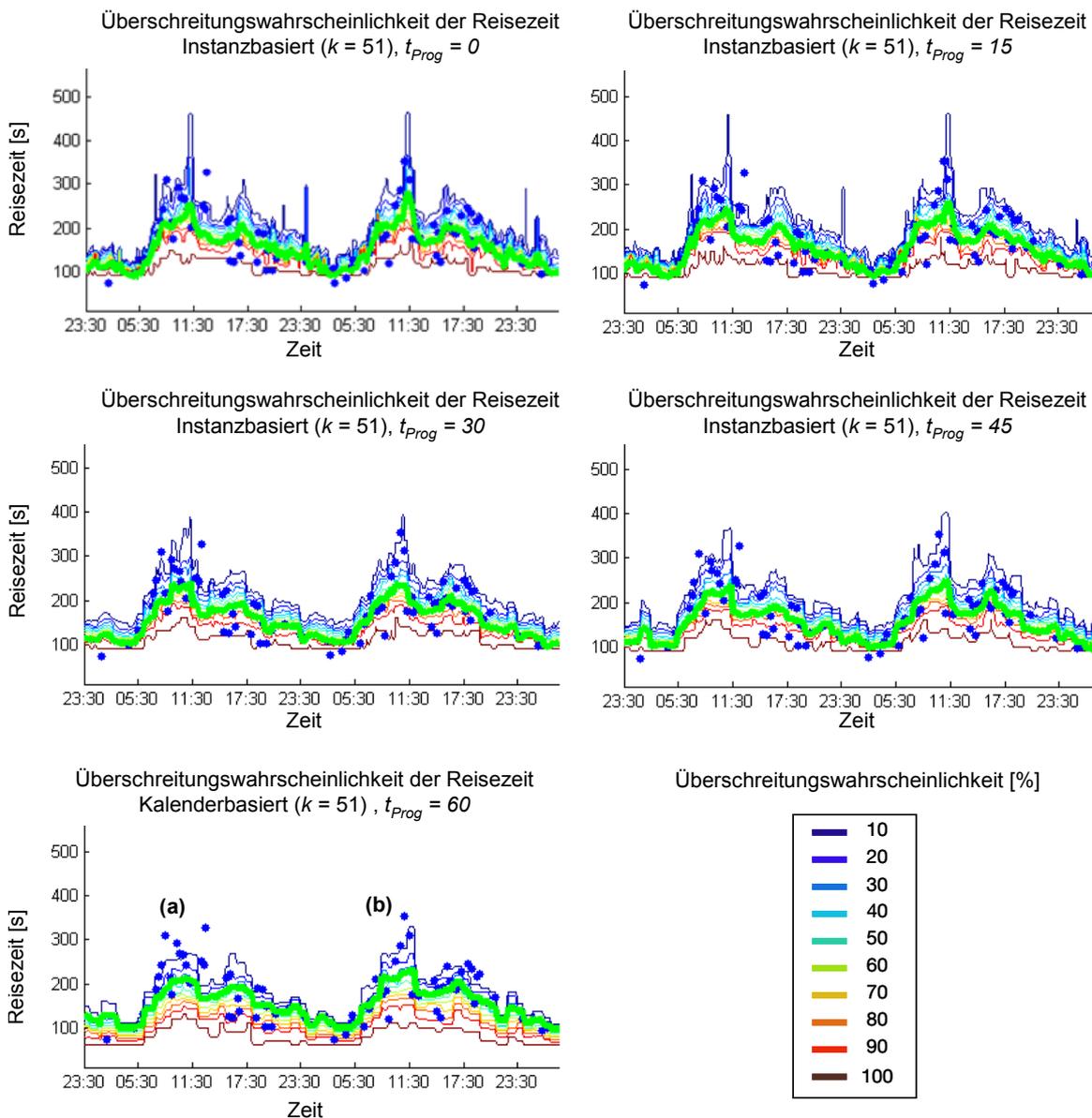


Abb. 5.27: Darstellung der Überschreitungswahrscheinlichkeit der Reisezeit über die Zeit für verschiedene Prognosehorizonte und kalenderbasiert auf dem "OPERNRING"

5.8 Fazit der Bewertung

Insgesamt zeigt sich, von den Ausreißern abgesehen, dass weder die kalenderbasierte noch die Trivialprognose eine Alternative zu komplexeren Verfahren unter Berücksichtigung von Verkehrsmustern darstellen. Es ist dabei relativ unbedeutend, ob nur die lokale Zeitreihe oder ein räumlich ausgedehntes Muster verwendet wird. Bzgl. der Wahl des Mustererkennungsverfahrens (Instanzbasiert oder Künstliches Neuronales Netz) zur Prognose von Erwartungswerten können keine eindeutigen Empfehlungen gegeben werden. Letztendlich hängt die Wahl von der Präferenz des Nutzers ab.

Da bei der Prognose des lokalen Belegungsgrades etwas bessere Ergebnisse mit den MLF erreicht wurden und da aufgrund der größeren Anzahl an Beobachtungen und eines entsprechend größeren Datenarchivs die Recheneffizienz ein Rolle spielen könnte stellen MLF hier eine gute Wahl dar.

Für die Schätzung und die Prognose von Reisezeiten bringt der Einsatz des Instanzbasierten Lernverfahrens insbesondere aufgrund der Fähigkeit, relativ unkompliziert Kenngrößen über den Erwartungswert hinaus zu ermitteln, Vorteile. Dies bildet gleichzeitig den größten Mehrwert des Verfahrens im Vergleich zu MLF, bei denen dies nicht ohne weiteres in dieser Art möglich ist.

6 Fazit und Ausblick

6.1 Zusammenfassung

Verfahren und Ergebnisse

Es wurde ein Instanzbasiertes Lernverfahren zur Prognose von verkehrlichen Kenngrößen konzipiert und zur Prognose lokaler Belegungsgrade und Reisezeiten eingesetzt. Dem Verfahren liegt die Idee zugrunde, zufällige Beobachtungen als abhängige Variable gemeinsam mit Mustern kontinuierlich verfügbarer Daten als unabhängige Variable zu archivieren. Zur Prognose können dann die aktuellen unabhängigen Variablen mit den historischen Mustern verglichen werden und auf Basis der k ähnlichsten Muster die Prognose berechnet werden.

Als unabhängige Variable wurden lokale Belegungsgrade bzw. lokale Verkehrsstärken, sowie Kalenderinformation verwendet. Als abhängige Variable dienten lokale Belegungsgrade und Taxireisezeiten (Graz) sowie Reisezeiten aus Fahrzeugwiedererkennung (München). Ein wichtiges Merkmal des Verfahrens ist, dass die unabhängigen Variablen nicht zwingend am betrachteten Detektor selber bzw. nicht auf der betrachteten Route erfasst werden müssen, sondern dass die Messwerte entsprechend ihrer Distanz von der betrachteten Lokation und dem Zeitversatz zwischen ihrem Bezugsintervall und dem aktuellen Zeitpunkt gewichtet eingehen.

Das Verfahren besitzt sechs Parameter: Parameter zur räumlichen (ds_{max}) und zeitlichen (dt_{max}) Gewichtung der einzelnen Merkmale innerhalb des Verkehrsmusters, den Parameter k_e zur überproportionalen Gewichtung stark differierender Merkmale im Verkehrsmuster, den prognosehorizontabhängigen Parameter G_V zur Gewichtung des Verkehrsmusters im Vergleich zu den Kalendermerkmalen (Tagesklasse und Tageszeit), die Anzahl der zu berücksichtigenden Instanzen k und den Parameter K_W zur Gewichtung der Instanzen bei der Prognose entsprechend ihrer Gesamtähnlichkeit.

Eine Analyse hat gezeigt, dass das Verfahren bezüglich der Parametereinstellungen sensitiv ist. Dies ist einerseits günstig, da das Verfahren an Randbedingungen angepasst werden kann und die Parameter ansonsten entfallen könnten (womit grundlegende Annahmen im Modellentwurf in Frage gestellt würden). Andererseits macht es eine Justierung der Parameter notwendig. Vergleichsweise geringen Einfluss hat dabei lediglich die Wahl von dt_{max} und k_e . Aufgrund zu erwartender Wechselwirkungen zwischen den Parametern wurde zur automatischen Kalibrierung ein Genetischer Algorithmus eingesetzt. Dieser optimiert die Parameter gleichzeitig und benötigt lediglich eine eindeutig definierte Zielfunktion als Optimierungskriterium (hier: Minimale Abweichung zwischen Prognose- und Referenzwert).

Mit den in den Optimierungsläufen gefundenen Parametern als Vorgabe wurde das Instanzbasierte Lernverfahren auf Basis realer Daten getestet und mit verschiedenen Verfahren verglichen. Als Vergleichsverfahren aus dem Bereich der Mustererkennung wurde eine MLF Topologie zur Verarbeitung der räumlich-zeitlichen Verkehrsmuster und der Kalendermerkmale entworfen und implementiert.

Bei der Prognose des lokalen Belegungsgrades, untersucht an zwei Detektoren in Graz, wurden gute Ergebnisse erzielt. "Gut" bedeutet, dass die aggregierten Fehlermaße relativ niedrig sind und dass eine visuelle Analyse zeigt, dass die Belegung gut abgebildet wird. Bzgl. der Fehlermaße liefert das Instanzbasierte Lernverfahren bei Horizonten bis 15 Minuten vergleichbare Ergebnisse wie das MLF und die Trivialprognose. Bei Prognosehorizonten kleiner als 60 Minuten wurden stets bessere Ergebnisse erreicht als mit der Trivialprognose und als auf Basis rein kalenderbasierter Information.

Die Belegung konnte auch ohne Berücksichtigung der aktuell an dem zur Prognose betrachteten Detektor erfassten Messwerte, also alleine aus der Information umliegender Detektoren, geschätzt und prognostiziert werden. Dies induziert, dass Verkehrsmuster netz- bzw. teilnetzweite Ausdehnungen haben, also Muster umliegender Detektoren Prädiktoreigenschaften besitzen, die etwa zur Messwertersetzung oder Plausibilisierung genutzt werden können.

Prognose und auch Schätzung der Reisezeit zeigten sich in beiden Anwendungsfällen erwartungsgemäß als herausfordernder als die Prognose der Belegungsgrade. Konkret heißt das, dass insbesondere in Bereichen hoher Reisezeiten eine Verbesserung der kalenderbasierten Schätzung oder Prognose des Erwartungswertes zwar schwierig ist, jedoch erreicht werden konnte. Dies betrifft sowohl das hier entwickelte Instanzbasierte Lernverfahren, als auch das MLF.

Ergänzungen des Instanzbasierten Lernverfahrens durch Heuristiken ermöglichen eine speziell an hohe Reisezeiten angepasste Schätzung und Prognose der Reisezeit. Die Ermittlung von Zuverlässigkeitsindikatoren und die Schätzung von Reisezeitverteilungen stellen nutzbringende Zusatzinformationen dar, die – je nach Nutzung – einen deutlichen Mehrwert über den Erwartungswert hinaus darstellen können.

Die in Kapitel 1 definierten Anforderungen dürfen als im Wesentlichen erfüllt angesehen werden. Aus funktionaler Sicht wurden Prognosehorizonte bis 60 Minuten realisiert, wobei die Historie abhängig vom Prognosehorizont gewichtet wird. Es wurden Zuverlässigkeitsindikatoren dargestellt und deren Wirkung exemplarisch gezeigt. Die Schätzung von erwarteten Verteilungen des Prognosewertes ist mit dem Ansatz ebenfalls möglich.

Das Verfahren arbeitet auf Basis realer, nach dem derzeitigen Stand der Technik verfügbarer Daten (Reisezeiten aus Taxipositionsmeldungen und Fahrzeugwiedererkennung, lokale Ver-

kehrdaten) und berücksichtigt dabei insbesondere, dass die Reisezeiten häufig nur zufällig oder zeitlich begrenzt beobachtet werden können. Das Modell arbeitet prinzipiell auf unterschiedlichen Datenquellen. In einem neuen Einsatzfeld muss "lediglich" spezifiziert werden, welche Daten als unabhängige Variable und welche als abhängige Variable verwendet werden sollen. Ferner müssen die geometrischen Randbedingungen hinterlegt werden. Die Modellkalibrierung ist ein nicht zu unterschätzender Arbeitsschritt. Automatische Methoden können eingesetzt werden, eine visuelle Analyse und ggf. Nachjustierung ist aber insbesondere bei der Reisezeitschätzung und -prognose ratsam.

Die Robustheit gegenüber fehlenden Daten ist explizit gegeben, da die Musterähnlichkeiten nur auf Basis der vorhandenen Daten berechnet werden und die entstehende Unsicherheit aufgrund fehlender Messwerte mit einem Strafterm eingehen. Das Instanzbasierte Verfahren ist per Definition adaptiv, da neue Datensätze als Instanzen einfach nur dem Archiv hinzugefügt werden müssen.

Insgesamt erscheint die vorgestellte Nutzung von räumlich-zeitlichen Verkehrsmustern zusammen mit Kalendermerkmalen als ein möglicher und sinnvoller Weg zur Prognose verkehrlicher Kenngrößen auf Basis aktueller und historischer Daten.

Mögliche Anwendungen im Verkehrsmanagement

Das entwickelte Verfahren kann in verschiedenen Bereichen des Verkehrsmanagements eingesetzt werden. Voraussetzung ist jeweils ein umfangreiches Archiv historischer Beobachtungen und die Erfassung von Echtzeitdaten. Typischerweise würde das Verfahren in einer Verkehrsmanagementzentrale eingesetzt werden. Beispielhaft seien im Folgenden einige Anwendungsszenarien für das Verfahren genannt.



Abb. 6.1 Verkehrsqualitätskarten für Graz

In Verkehrsinformationsportalen können die mittels des Verfahrens prognostizierten Reisezeiten und lokalen Kenngrößen übersetzt in eine Verkehrsqualitätsstufe für einen Zeitpunkt oder ein Zeitintervall (z.B. "für die nächste viertel Stunde") farbkodiert auf dem Netz dargestellt werden. Üblicherweise werden die Farben rot, gelb und grün verwendet und stellen dabei i.d.R. jeweils die Zustände "gestaut", "dicht" und "frei" dar. Neben der Ermittlung der relevanten verkehrlichen Kenngrößen ist daher noch eine Logik erforderlich, die den Kenngrößen einen qualitativen Verkehrszustand und somit eine Farbe zuweist. Da die Interpretation der Farben einer Verkehrsqualitätskarte stark vom Empfinden des Betrachters abhängt, sollten die zugrunde liegenden Schwellenwerte distanzneutral (also als mittlere Geschwindigkeit statt als Reisezeit) angegeben werden. Trotz möglicher Fehlinterpretationen geben netzhafte Darstellungen einen schnellen Überblick und sind heute in der Praxis weit verbreitet. Im Rahmen des IVM Graz werden Verkehrsqualitätskarten generiert, wobei eine 15-Minuten-Prognose der Reisezeiten auf Basis von Taxipositionsmeldungen als Grundlage verwendet wird (Beispiele siehe Abb. 6.1). Soweit vorhanden, wird die Darstellung bei Abweichungen von den lokalen Belegungsgraden überschrieben ("Fusion durch Überlagerung"), da diese aufgrund der größeren "Mess- und Prognosesicherheit" als zuverlässiger betrachtet werden.

Die mit dem Verfahren abschätzbaren Verteilungen der zu erwartenden Reisezeit (übertragbar auf lokale Verkehrskenngrößen und abgeleitete Zustände) können als Eingangswerte für die sogenannte zuverlässige dynamische Navigation verwendet werden. Die zuverlässige dynamische Navigation zeichnet sich im Gegensatz zur "normalen" dynamischen Navigation durch die explizite Berücksichtigung der zu erwartenden Zuverlässigkeit einer Route aus. KAPARIAS ET AL. [2007] stellt mit dem Verfahren und dem Programm ICNavS (Imperial College Navigation Software) eine Methode zur zuverlässigen Zielführung vor. Dabei wird im Prinzip nicht primär nach der schnellsten Route (also der mit dem niedrigsten Erwartungswert), sondern nach derjenigen Route mit der niedrigsten Störungswahrscheinlichkeit bei einer maximal zulässigen erwarteten Reisezeit als Nebenbedingung gesucht.

Das Verfahren erlaubt die Schätzung von Detektormesswerten auf Basis historischer Zusammenhänge in Echtzeit alleine aus Messungen stromaufwärtiger Detektoren. Dadurch können fehlende Messwerte ersetzt werden und bei gleichzeitiger Messung und Schätzung der Messwerte die Messwerte plausibilisiert werden. Bei der Plausibilisierung würde davon ausgegangen, dass unter "normalen" Bedingungen der Messwert aus den umliegenden Detektoren geschätzt werden kann und bei einer Abweichung zwischen Schätzung und Messung entweder eine untypische Situation vorliegt oder der betrachtete Detektor falsche Messwerte liefert. Selbst wenn eine automatische Entscheidung, ob es sich um eine Fehlmessung oder eine außergewöhnliche Situation handelt, schwierig oder gar nicht möglich ist, wäre eine Benachrichtigung des Bedienpersonals in jedem Fall sinnvoll.

6.2 Möglichkeiten zur weiteren Entwicklung

Erweiterung des Verfahrens

Zur Archivierung als Instanzen wurden die Reisezeiten eines Zeitintervalls (15 Minuten) zum Median aggregiert. Vor dem Hintergrund, dass auch hier möglicherweise Informationen verloren gehen und dass das Instanzbasierte Lernen an sich eine glättende Eigenschaft hat, wäre es zu untersuchen ob eine Archivierung als Einzelfahrzeugreisezeit sinnvoller wäre. Es könnten dann mehrere Einzelfahrzeugreisezeiten demselben Muster von unabhängigen Variablen zugeordnet werden, so dass die k nächsten Nachbarn theoretisch aus einem Zeitintervall gestellt werden könnten und nicht zwingend aus k Intervallen. Die Aggregation auf 15 Minuten-Intervalle hat sich bewährt und erscheint als guter Kompromiss zwischen Aktualität und Rauschunterdrückung. Je nach Kontext sollten jedoch kürzere Intervalle verwendet und untersucht werden (z.B. außerorts, Prognose von lokalen Geschwindigkeitseinbrüchen).

Die in Abschnitt 4.5 vorgestellten Zuverlässigkeitskonzepte versprechen Potential und Mehrwert für praktische Anwendungen. Jedoch sind sie trotz Nutzung statistischer Konzepte auch intuitiver Natur und könnten weiter verfeinert werden. Insbesondere eine Analyse der Prädiktorqualitäten der Zuverlässigkeitsindikatoren in verschiedenen Verkehrszuständen und die Fusion der Indikatoren zur Beurteilung der "Gesamtqualität" einer Prognose dürfte ein sinnvoller Forschungsgegenstand sein.

Implementierungsaspekte

Ein wichtiger Forschungsgegenstand im Bereich der Instanzbasierten Lernverfahren und der Mustererkennung im Allgemeinen ist die informationsverlustfreie Reduktion der Wissensbasen. Wissensbasen können über die Zeit sehr groß werden, dementsprechend wachsen Speicherplatzbedarf und die benötigte Zeit zum Durchsuchen der Datenbasis bei Instanzbasierten Lernverfahren. Eine Möglichkeit dem entgegen zu wirken ist die Reduktion der Größe der Datenbasis. Dazu können komplette Merkmalsvektoren entfernt bzw. zusammengefasst werden oder die Anzahl der Merkmale je Muster reduziert werden. Grundidee ist jeweils die Entfernung von Redundanzen in der Datenbasis. Beispielsweise können gleiche oder sehr ähnliche Merkmalsvektoren mit Clusteralgorithmen, also unüberwacht arbeitenden Verfahren, zu Repräsentanten zusammengefasst werden. Dabei könnte sich ein hierarchisches Verfahren als günstiger erweisen als ein partitionierendes Verfahren. Hierarchische Verfahren erstellen eine Baumstruktur, wobei die Schichten unterschiedliche Aggregierungsstufen darstellen und somit die Anzahl der Cluster nicht vorgegeben werden muss. Auf Musterebene kann versucht werden, die Dimension der Muster zu reduzieren. Ein gängiges Verfahren dazu ist die Hauptkomponentenanalyse (*engl.: Principle Component Analysis, PCA*), mittels der Faktoren als Linearkombinationen der einzelnen Merkmale gebildet werden. Der Infor-

mationsgehalt der Faktoren kann über die sogenannten Eigenwerte bestimmt werden, so dass redundante Information identifiziert und exkludiert werden kann.

Bei der Datenreduktion muss vorsichtig vorgegangen und Nachteile müssen in Kauf genommen werden. Zum einen geht immer Information verloren. Insbesondere wird die Konstruktion von Vertrauensintervallen für die Prognose basierend auf der aktuellen Situation schwierig. Auf Musterebene müssen die Reduktionsverfahren auch mit unterschiedlich vollständigen Mustern und einzelnen fehlenden Messwerte umgehen können, damit sie durchgängig angewendet werden können.

Effizienzsteigernde Maßnahmen, wie beispielsweise die Reorganisation der Daten als kd (= *k-dimensional*) Baum sind prinzipiell eine Option. Allerdings müssen hier die Gewichte bei der Ähnlichkeitsberechnung speziell berücksichtigt werden. Darüber hinaus muss das entstehende Konstrukt weiterhin robust gegenüber fehlenden Merkmalen sein.

Ergänzung um weitere Datenquellen und Datenvorverarbeitung

Das vorgestellte Verfahren arbeitet als rein datengetriebenes Verfahren im Wesentlichen unabhängig von der Art der Eingangsdaten. Entscheidend ist, dass ein Zusammenhang zwischen unabhängigen und abhängigen Variablen besteht und dass die Daten entsprechend strukturiert sind. Die Verwendung weiterer Datenquellen ist somit technisch und funktional möglich.

Neben Reisezeiten aus Taxipositionsmeldungen und Fahrzeugwiedererkennung existieren weitere interessante Datenquellen zur Beobachtung streckenbezogener Verkehrskenngrößen. Innerorts von großer praktischer Relevanz wäre die Nutzung von Fahrzeugpositionsmeldungen des straßengebundenen ÖPNV zur Reisezeiterhebung und zum Einsatz als abhängige Variable. Diese besitzen typischerweise eine hohe Verlässlichkeit und können "planbar", entsprechend dem Fahrplan, erhoben werden. Um diese Daten als Informationsgrundlage für den IV nutzen zu können, müssen die Daten von ÖPNV-typischen Verlustzeiten bereinigt werden.

Zur Nutzung als unabhängige Variable könnten die LSA-Signalisierungszustände eine wichtige Rolle spielen. Insbesondere könnten dann auch Verkehrsstärkemessungen besser genutzt werden. Beispielsweise könnten kombinierte Kenngrößen wie der "Auslastungsgrad" oder die "Dauer der Überlastung" als beschreibende Variable zur verbesserten Beschreibung der Verkehrssituation eingeführt werden.

Außerorts stellen Luftbilder, z.B. aus Befliegungen eine weitere Datenquelle mit hohem Informationsgehalt dar. Aus Aufwands- und Kostengründen werden derartige Daten jedoch selten erhoben. Interessant wäre die Anlage eines Archivs mit streckenbezogenen Dichte-, Fluss- und Geschwindigkeitsfeldern und zugehörigen lokalen Daten. Ab einer gewissen, zu

eruierenden Stichprobengröße lassen sich möglicherweise typische Muster identifizieren, die zur Schätzung und zur Kurzfristprognose der Verkehrssituation zwischen den Messquerschnitten genutzt werden können. Des Weiteren stellen außerorts lokal erhobene Daten aufgrund der hohen Erfassungsdichte selbst eine für die Kurzfristprognose interessante Quelle dar. Die zuverlässige Prognose von Geschwindigkeitseinbrüchen unter Nutzung stromabwärtiger und stromaufwärtiger Information würde neben einer frühzeitigen Warnung auch ohne die explizite Nutzung eines Verkehrsflussmodells einen Einblick in die Verkehrs- bzw. Staudynamik erlauben.

Schließlich können auch aggregierte Kenngrößen, beispielsweise Verkehrsqualitätsstufen, dem Verfahren als abhängige Variable präsentiert und direkt prognostiziert werden. Aus dem Regressionsproblem würde dann ein Klassifikationsproblem, so dass geeignete Auswahlmechanismen (etwa "Mehrheitsentscheid") an Stelle der lokalen Approximationsfunktion treten müssten.

Kopplung mit weiteren Modellen

Das vorgestellte Verfahren liefert in den meisten Situationen sowohl für die Schätzung als auch für die Prognose von lokalen Verkehrskenngrößen und Reisezeiten gute Ergebnisse. Ein Instanzbasiertes Lernverfahren kann per Definition jedoch nur dann gut funktionieren, wenn mindestens ein dem aktuellen Muster ähnliches Muster bereits beobachtet wurde und als Instanz vorliegt. Auch wenn die Datenbasis im Laufe der Zeit immer vollständiger wird, können unbekannte Situationen auftreten. Es erscheint daher günstig, das vorgestellte Verfahren mit Modellen zu koppeln, die auch in unbekanntem Situationen sinnvolle Prognosen liefern können.

Die Kopplung mit einem Verkehrsablaufmodell kann einen Mehrwert zur Prognose in zum Zeitpunkt der Prognose unbekanntem und somit nicht archivierten Situationen bringen. Dabei bezieht sich "unbekannt" auf bis dato nicht beobachtete Angebot-Nachfrage-Kombinationen, wie beispielsweise die Sperrung einer Straße oder eine erstmalig auftretende Nachfragespitze aufgrund eines neuen Attraktors. In diesen Situationen kann von einem Musterkennungs-basierten Verfahren alleine keine zuverlässige Prognose geleistet werden. In einem Verkehrsablaufmodell bzw. einem Verkehrssimulationsmodell wird das Angebot explizit modelliert, entsprechend können die verkehrlichen Wirkungen bei Angebotsänderungen abgeschätzt werden. Auch gemessene untypische Nachfragen können simuliert werden. Zur Prognose untypischer Nachfragen sind allerdings weitere Zusatzinformationen – etwa das Attraktionspotential eines neuen Einkaufszentrums – nötig.

Als weitere Ergänzung bzw. als übergeordnete Instanz wäre ein Expertensystem geeignet. Expertensysteme beruhen im Gegensatz zu rein datenbasierten Ansätzen auf verallgemeinerten Regeln. Diese können so angelegt werden, dass sie über den gesamten möglichen Merkmalsraum des Eingabemusters sinnvolle Ergebnisse liefern. Die Regeln in Expertensys-

temen orientieren sich naturgemäß an dem bestehenden Datenarchiv, da dieses üblicherweise das Wissen enthält, das auch dem Experten (meist Verkehrsingenieur) zur Verfügung steht. Darüber hinaus können die Regeln entsprechend dem Expertenwissen in den Merkmalsraum "extrapoliert" werden. Beispielsweise könnte die Reisezeit als Funktional der Belegungsgrade außerhalb des beobachteten Wertebereichs bei steigenden Belegungsgraden asymptotisch oder linear gegen einen vorzugebenden Maximalwert gehen. Gekoppelt mit einem entsprechenden Hinweis, dass dieser Wert nicht auf einer Beobachtung sondern einer Annahme beruht, wäre auch dies eine nützliche Information.

Notation und Definitionen

Symbole, Variablen und Parameter

η	Angleichfaktor bei der Prognose unter Berücksichtigung des aktuellen Messwertes
ρ	Lokale Verkehrsdichte [Fz/km]
Δt	Betrachtungs- / Aggregierungsintervall der verkehrlichen Kenngrößen
σ_y	gewichtete Varianz der Stichprobe $y_{historisch}$
b	Lokaler Belegungsgrad [%]
D	Distanzmaß zur Beschreibung der Unähnlichkeit zweier Merkmalsvektoren
D_K	Distanzmaß zur Beschreibung der Unähnlichkeit zweier Kalendermerkmale
ds_{max}	Parameter zum Einstellen des Einflusses von Detektoren entsprechend ihres Abstands zur Lokation der abhängigen Variablen.
$ds_{r,dr}$	Abstand von Lokation r nach dr im Netz (stromaufwärts von r aus gesehen)
$dt_{i,j}$	Zeitdifferenz zwischen dem aktuellen und dem zu Merkmal j der Instanz i gehörendem Zeitintervall
dt_{max}	Parameter zum Einstellen des Einflusses von Messwerten entsprechend ihres zeitlichen Abstandes zum aktuellen Zeitintervall.
$D_{V,i}$	Distanzmaß zur Beschreibung der Unähnlichkeit des aktuellen Verkehrsmusters und des Historischen Verkehrsmusters i
e_T	Abweichung mit erlaubter Toleranz ΔT
g_j	Faktor zur Gewichtung des Verkehrsmerkmals j in Abhängigkeit seiner räumlichen und zeitlichen Distanz zum betrachteten abhängigen Variablen y . ($g=f(ds,dt,ds_{max},dt_{max})$)
G_V	Parameter zur Gewichtung des Einflusses des Verkehrsmusters
i	Laufindex, i.d.R. zur Bezeichnung einer Instanz ($i = 1 \dots \text{Anzahl der Instanzen}$)
j	Laufindex, i.d.R. zur Bezeichnung eines Verkehrsmerkmals innerhalb einer Instanz/eines Musters ($j = 1 \dots \text{Anzahl der Merkmale je Instanz}$)
k	Anzahl der beim Instanzbasierten Lernen verwendeten Instanzen ("Nachbarn")
k_e	Exponent der Minkowski Metrik zur (über)proportionalen Gewichtung ($k_e \geq 1$) großer Distanzen innerhalb eines Verkehrsmusters
K_w	Kernweite der lokalen Gewichtungsfunktion
L_x	Lage eines Merkmalsvektors zu den k nächsten historischen Instanzen
m	Anzahl der Merkmale einer Instanz / eines Musters
n	Anzahl der Prognose verwendbaren archivierten Instanzen / Muster
$n_{gesamt,i}$	Anzahl der Merkmale der Instanz i

$n_{gültig,i}$	Anzahl der gültigen Merkmale der Instanz i
P	Parameter
p	Laufindex zur Beschreibung einer realisierten Einzelprognose, $p = 1 \dots P$
ρ	Wahrscheinlichkeit
PZ	Anzahl der realisierten Einzelprognosen
q	Lokale Verkehrsstärke [Fz/h]
r	Laufindex für Routen
R	Range, einfaches statistisches Streuungsmaß für Verteilungen
r^2	Bestimmtheitsmaß
$r_{yy'}$	Korrelationskoeffizient für Datenreihen y und y'
s	Empirische Standardabweichung
S_F	Strafterm für fehlende Messwerte
S_x	Distanz der zur Prognose verwendeten Instanzen zum aktuellen Muster
t	Index für das Zeitintervall $[t, t + \Delta t]$
T_i	Tagesklassenmerkmal der Instanz i
t_{netto}	Nettozeitlücke zwischen zwei Fahrzeugen [s]
t_{prog}	Prognosehorizont [min]
TTC	Time to Collision [s], verbleibende Zeit bis zu einer Kollision
TT_r^i	Einzelfahrzeugreisezeit [s] des Fahrzeugs i auf der Route r
$TT_r^{j,t, \Delta t+t}$	Mittlere Reisezeit [s] auf einer Route bezogen auf das Zeitintervall $[t, t + \Delta t]$
TZ	Merkmal Tageszeit
v	Lokale Geschwindigkeit [km/h]
Var / s^2	Empirische Varianz
w_i	Faktor zur Gewichtung der Instanz i in Abhängigkeit ihrer Distanz D_i ($w_i = f(D_i)$)
$X_{aktuell}$	Aktueller Merkmalsvektor
X_i	Merkmalsvektor i , bestehend aus Merkmalen $x_{ij}, j = 1, \dots, n$
x_j	Unabhängige Variable j , "Merkmal j "
y'	Schätzwert für die abhängige Variable y
$y'_{angepasst}$	Schätzwert für die abhängige Variable y , an den aktuellen Messwert angepasst
y_i	Abhängige Variable, korrespondierend zum Merkmalsvektor i
Z_i	Zeitmerkmal der Instanz i
α	Gewichtungsfaktor bei der exponentiellen Glättung

Begriffe

<i>Instanz</i>	Eintrag im Datenarchiv, bestehend aus verkehrlichen und kalenderbasierten Merkmalen (unabhängige Variable) X_i und der Zielgröße y_i (abhängige Variable).
<i>Kante</i>	Eine Kante (<i>engl.: Link</i>) ist eine gerichtete Verbindung zwischen zwei Knotenpunkten. Sie enthält als Attribute die Anzahl der Fahrstreifen und eine Länge.
<i>Klassifikation</i>	Die Merkmalsvektoren werden in ein vorab festgelegtes Klassenschema (z.B. "Stau", "dicht", "frei") eingeordnet. Klassifikationsverfahren lernen überwacht anhand von Eingabe-Ausgabe Trainingsmustern die gegebenen Merkmalsvektoren korrekt zuzuordnen.
<i>Messquerschnitt</i>	Ein Messquerschnitt ist ein Ort auf einer Kante, dem lokale Verkehrskenngrößen zugeordnet sind. Wird durch die Zusammenfassung mehrerer (mindestens einer) Messstelle gebildet.
<i>Messstelle</i>	Ort auf einer Kante und dort auf einem Fahrstreifen, an dem lokale Verkehrskenngrößen erfasst werden. Spezifiziert durch die ID der Kante, den Offset (Abstand vom Anfang der Kante) und den Fahrstreifen.
<i>Prognose</i>	Aussage über Zustände in der Zukunft, Schätzung für zukünftige Zeitpunkte.
<i>Regression</i>	Die Merkmalsvektoren werden – im Gegensatz zur Klassifikation – auf ein Zahlenspektrum abgebildet. Anhand von Eingabe-Ausgabe Trainingsmustern lernt das Modell überwacht, den Merkmalsvektor als unabhängige Variable auf einen Zielwert (abhängige Variable) abzubilden.
<i>Route</i>	Eine Route (auch Strecke) ist eine Zusammensetzung mehrerer Kanten, i.d.R. ein sinnvoller Weg durch das Netz.
<i>Schätzung</i>	Näherungsweise Bestimmung von Zahlenwerten, u.a. durch statistisch-mathematische Methoden. Das Ergebnis einer Schätzung weicht im Regelfall vom wahren Wert ab
<i>Segmentation</i> (<i>engl.: Clustering</i>)	Die gegebenen Merkmalsvektoren einer Datenbasis werden automatisch entsprechend vorgegebener Abstandsmaße optimal segmentiert, also in Gruppen (Cluster) eingeteilt. Da keine Klassen vorgegeben sind, lernen Segmentationsverfahren unüberwacht.

Abkürzungen

<i>ANFIS</i>	Adaptive-Network-Based Fuzzy Inference System
<i>ANN</i>	Artificial Neural Network, künstliches neuronales Netz
<i>ANPR</i>	Automatic Number Plate Recognition, Automatische Nummernschilderkennung
<i>AR(I)MA(X)</i>	Auto Regressive (Integrated) Moving Average (with eXternal)
<i>ARCH</i>	Autoregressive Conditional Heteroscedastic
<i>BLIDS</i>	Bluetooth Based Traffic Data Collection System
<i>CAN</i>	Controller Area Network
<i>ESP</i>	Elektronisches Stabilitätsprogramm für Kraftfahrzeuge
<i>EVS</i>	Erweiterte Vorwärts Stern (Struktur)
<i>FCD</i>	Floating Car Data, fahrzeuggenerierte Daten
<i>FLEET</i>	Fleet Logistics Service Enhancement with Egnos & Galileo Satellite Technology

<i>GA</i>	Genetischer Algorithmus
<i>GAOT</i>	Genetic Algorithm Optimization Toolbox
<i>GL</i>	Ganglinie
<i>GPS</i>	Global Positioning System
<i>IL</i>	Instanzbasiertes Lernen
<i>IVM</i>	Integriertes Verkehrsmanagement
<i>kNN</i>	K Nearest Neighbour, k Nächste Nachbarn
<i>KNN</i>	Künstliches Neuronales Netz
<i>LOS</i>	Level of Service, Qualitätsstufe des Verkehrsablaufs
<i>LSA/LZA</i>	Lichtsignalanlage / Lichtzeichenanlage
<i>LWR</i>	Locally Weighted Regression, lokal gewichtete Regression
<i>MAF</i>	Mittlerer Absoluter Fehler
<i>MAVE</i>	Mobile Anlage zur Verkehrsdatenerfassung
<i>MAXF</i>	Maximaler Fehler
<i>MLF</i>	MultiLayer Feedforward Netz (spezielle Topologie eines ANN)
<i>MOTIV</i>	Mobilität und Transport im intermodalen Verkehr
<i>MPF</i>	Mittlerer Proportionaler Fehler
<i>MQ</i>	Messquerschnitt
<i>MQF</i>	Mittlerer Quadratischer Fehler
<i>MRF</i>	Mittlerer Relativer Fehler
<i>NOSTRADAMUS</i>	New Online Service for Traffic Data Multiplexing
<i>OCR</i>	Optical Character Recognition
<i>OnVLS</i>	Online Verkehrslagesystem
<i>ÖPNV</i>	Öffentlicher Personen Nahverkehr
<i>PDF</i>	Probability Density Function, Wahrscheinlichkeitsdichtefunktion
<i>PFE</i>	Path Flow Estimator
<i>RFID</i>	Radio Frequency Identification
<i>SCOOT</i>	Split Cycle Offset Optimisation Technique
<i>SIM</i>	Subscriber Identity Module
<i>STARIMA</i>	Space Time ARIMA
<i>TMC</i>	Traffic Message Channel
<i>V2I</i>	Vehicle to Infrastructure, Kommunikation zwischen Fahrzeug und Infrastruktur
<i>VISUM</i>	Verkehr in Städten - Umlegung
<i>VMZ</i>	Verkehrsmanagementzentrale

<i>WLAN</i>	Wireless Local Area Network
<i>xFCD</i>	Extended Floating Car Data, erweiterte fahzeuggenerierte Daten
<i>XML</i>	Extensible Markup Language

Quellenverzeichnis

- ABDEL-ATY, M.; KITAMURA, R.; JOVANIS P. [1996]: Investigating Effect of Travel Time Variability on Route Choice Using Repeated Measurement Stated Preference Data. *Transportation Research Record*, 1493, pp. 39-45.
- AUSSENEGG, W.; MIAZHYNKAIA, T. [2006]: Uncertainty in Value-at-Risk Estimates under Parametric and Non-parametric Modeling, "Topics in Banking and Finance – Riskmanagement", WU-Wien, Juni 2006.
- BAJWA, S.; CHUNG, E.; KUWAHARA, M. [2003]: Sensitivity Analysis of Short-Term Travel Time Prediction Model's Parameters.
- BELL, M. G. H.; SHIELD, C. M., BUSCH, F.; KRUSE, G. [1997]: A Stochastic User Equilibrium Path Flow Estimator. *Transportation Research, Part C, Vol. 5 (1997), No. 3 / 4*, p. 197-210.
- BELZNER, H. [2002]: Ein hybrides Modell zur Kurzfristprognose lokaler Verkehrskenngrößen. Diplomarbeit an der Fakultät für Mathematik der TU München. Februar 2002.
- BENZ, T.; HOYER, R.; KATES, R.; MATHIAS, P.; MÜCK, J.; REHBORN, H.; SCHELL, B. [2003]: Information Supply for Intelligent Routing Services – The INVENT Traffic Network Equalizer Approach, 2003.
- BLATNIG, S.; SCHELANDER, M.: Extending the Information Horizon through Floating Car Data (<http://vi.uni-klu.ac.at/publications/studentwork/4.pdf>, Stand: 06/2008).
- BOX, G.E.P; JENKINS, G.M. [1970]: *Time Series Analysis, Forecasting and Control*, San Francisco, 1970.
- BRAXMEIER, H.; SCHMIDT, V.; SPODAREV, E. [2005]: Statistische Raum-Zeit-Analyse und Simulation von Verkehrsströmen in Ballungsgebieten. *Verkehrsforschung Online*, Beitrag 3, Ausgabe 2 – 2005.
- BÖHNKE, P. [1980]: Ein Verfahren zur aktuellen Analyse und Prognose der Verkehrsabläufe innerhalb von Straßenabschnitten aus lokal gemessenen Fahrzeugfolgen.
- BOGENBERGER, K. [2001]: Adaptive Fuzzy Systems for Traffic Responsive and Coordinated Ramp Metering. Dissertation am Fachgebiet Verkehrstechnik und Verkehrsplanung der TU München 2001.
- BRAUN, R., WEICHENMEIER, F. [2005]: Automatic offline-optimization of coordinated traffic signal control in urban networks using genetic algorithms. *Proceedings, 12th World Congress on Intelligent Transport Systems*, San Francisco, 06.-10. Nov. 2005.
- BREITENSTEIN, J. [1978]: Analyse und Weiterverfolgung gemessener Zeitreihen als Entscheidungsgrundlage für Verkehrsbeeinflussungsmaßnahmen. *Forschung Straßenbau und Straßenverkehrstechnik*, Heft 247, 1978.
- BUSCH, F.; HANITZSCH, A.; SCHENDZIELORZ, T. [2006a]: Prognose Verkehrstelematik 2015+ - Analysen, Vergleiche, Visionen. Im Auftrag der Robert Bosch AG, Hildesheim.
- BUSCH, F.; DINKEL, A.; LEONHARDT, A.; ZIEGLER, J.; KIRSCHFINK, H.; PETERS, J. [2006b]: Benchmarking für Verkehrsdatenerfassungs- und Verkehrssteuerungssysteme. Herausgegeben vom Bundesministerium für Verkehr, Bau und Stadtentwicklung, Heft 949, Bonn, ISBN 3-86509-575-5.
- CASSIDY, M. J.; WINDOVER, J. R. [1995]: Methodology for Assessing Dynamics of Freeway Traffic Flow. *Transportation Research Record Issue Number: 1484*, 1995.
- CEYLAN, H., BELL, M.G.H. [2004]: Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. *Transportation Research Part B* 38 (2004) 329-342.

CHEN, A.; ZHAOWANG, J.; RECKER, W. [2001]: Travel Time Reliability with Risk-Sensitive Travellers, UCI-ITS-WP-01-9.

CHEN, A.; RECKER, W. [2001]: Considering Risk Taking Behaviour in Travel Time Reliability. Paper presented at the Matsuyama Workshop on Transport Network Analysis, August 2000, Matsuyama, Japan, and at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C.

CHEN, Y., BELL M.G.H., WANG, D. [2005]: Risk-Averse time dependent route guidance by a constrained dynamic A* search in decenterized structure. Paper is submitted to the Transportation Research Board 2006, 2005-11-15.

CHEN, M.; CHIEN, S. [2001]: Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based. Transportation Research Board, 80th Annual Meeting, January 7-11, 2001 Washington, DC.

CHROBOK, R. [2005]: Theory and Application of Advanced Forecast Methods. Dissertation am Fachbereich Physik der Universität Duisburg-Essen, 2005.

CHU, L.; OH, J.; RECKER, W. [2005]: Adaptive Kalman Filter Based Freeway Travel time Estimation. 2005 TRB Annual Meeting.

CLARK, S. [2003]: Traffic Prediction using multivariate nonparametric regression. Journal of Transportation Engineering 129(2), 161–168.

COIFMAN, B. [1998]: Vehicle reidentification and travel time measurement in real-time on freeways using the existing loop detector infrastructure, Transportation Research Record, vol. 1643, pp. 181-191.

COIFMAN, B. [1999]: Identifying the Onset of Congestion Rapidly with existing Traffic Detectors. California PATH Working Paper UCB-ITS-PWP-99-17, November 1999.

COIFMAN, B. [2001]: A New Algorithm for vehicle Reidentification and Travel Time Measurement on Freeways.

COIFMAN, B. [2002]: Estimating travel times and vehicle trajectories on freeways using dual loop detectors, Transportation Research Part A, vol. 36, no. 4, pp. 351-364.

COVER, T. M.; HART, P. E. [1967]: Nearest Neighbor Patter Classification, IEEE Transactions on Information Theory 13(1), 21–27.

CREMER, M. [1995]: On the Calculation of Individual Travel Times by Macroscopic Models, 6th VNIS Con. & Pacific Rim Trans Tech. Conf., Seattle, July 1995.

DAGANZO, C. [1997]: Fundamentals of transportation and traffic operations. Pergamon, Oxford, ISBN 0-08-042785-5.

DAGANZO, C.; LAVAL, J. [2005]: Moving bottlenecks: A numerical method that converges in flow. Transportation Research Part B 39 (2005) 855-863.

DANECH-PAJOUH, M.; ARON, M. [1991]: ATHENA – a method for short-term inter-urban motorway traffic forecasting. Recherche Transports Securite–English Issue–No. 6, 1991.

DAVIS, G. A.; NIHAN, N. L. [1991]: Nonparametric regression and short-term freeway traffic forecasting, Journal of Transportation Engineering, 117(2), 178-188.

DEMIRÖZ, G.; GÜVENİR, H. A. [1996]: Genetic Algorithms to learn Feature Weights for the Nearest Neighbor Algorithm. Proceedings of BENELEARN-96 (1996) 117-126.

DIJKSTRA, E. W. [1959]: A note on two problems in connexion with graphs. Numerische Mathematik 1 (1959), S. 269–271.

- DION, F.; RAKHA, H. [2006]: Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. Article in press, Transportation Research Part B vol. 40, no. 9, pp. 745-766.
- EBERSPÄCHER, M. [1999]: Prognose der Eigenschaften stochastischer Prozesse mittels Neuro-naler Netze mit spezifischen Anwendungen in der Kommunikationstechnik: Dissertation am Insti-tut für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart, 1999.
- EHLERT, A.; BELL, M.; GROSSO, S. [2005]: The Optimisation of traffic count locations in road net-works. Transportation Research B Vol. 40, Issue 6, pp. 460-479, 2005.
- ESSER, J.; NEUBERT, L.; WAHLE, J.; SCHRECKENBERG, M. [1999]: Microscopic Online Simulation of Urban Traffic. Proceedings of the 14th ITS Worldcongress (Pergamon 1999), p. 535-554.
- FELLENDORF, M.; NÖKEL, K.; HANDKE, N. [2000]: VISUM-online – traffic management for the EXPO 2000 based on a traffic model. Traffic Technology International, Annual 2000.
- FERRARI, P. [1988]: The control of motorway reliability. Transportation Research A, 25, S. 419-427.
- FGSV [2005]:Hinweise zum Fundamentaldiagramm – Grundlagen und Anwendungen. FGSV, Arbeitsgruppe Verkehrsführung und Verkehrssicherheit, 2005.
- FÖRSTER, G. [2008]: Kurzfristprognose auf Basis von Raum-Zeit-Mustern. Forschungsgesell-schaft für Straßen- und Verkehrswesen (Hrsg.), HEUREKA '08, Optimierung in Verkehr und Transport - Tagungsband, FGSV Verlag, Köln, ISBN 978-3-939715-48-1.
- FORSTER, M. [1999]: Optimierung Künstlicher Neuronaler Netze mit genetischen Algorithmen. Diplomarbeit an der Universität Leipzig, 1999.
- FREUDENBERGER, P. [2001]: Analyse von Schleifendetektordaten und Entwicklung von Methoden zur Plausibilitätsprüfung. Diplomarbeit am Fachgebiet Verkehrstechnik und Verkehrsplanung der TU München, 2001.
- FRIEDRICH, B. [2004]: Ermittlung eines räumlich-zeitlich konsistenten Verkehrslagebildes. Stand der Technik und neue Entwicklungen. Universität Hannover, Institut für Verkehrswirtschaft, Straßenwesen und Städtebau, 2004.
- GAO, S.; CHABINI, I. [2006]: Optimal routing problems in stochastic time dependent networks. Transportation Research Part B 40 (2006) 93-122.
- GERDES, I.; KLAWONN, F.; KRUSE, R. [2004]: Evolutionäre Algorithmen, Vieweg Verlag. ISBN 3-528-05570-7.
- GÜHNEMANN, A. [2005]: New Approaches to Traffic Monitoring by "Floating Car Data", ITS Semi-nars October 1st 2003, DLR.
- GOODRICH, M.; TAMASSIA, R. [2001]: Algorithm Design: Foundations, Analysis and Internet Exam-ples. John Wiley & Sons, Inc. ISBN 0-471-38365-1, 2001.
- HAGEDORN, J.; BISSANTZ, N.; MERTENS, P. [1997]: Data Mining (Datenmustererkennung): Stand der Forschung und Entwicklung. Wirtschaftsinformatik 39 (1997), S. 601-612.
- HANDLEY, S.; LANGLEY, P.; RAUSCHER, F. [1998]: Learning to predict the duration of an automobile trip', Proceedings of the International Conference on Knowledge Discovery and Data Mining (1998, New York), S. 219-223.
- HELBING, D. [1997]: Verkehrsdynamik: Neue physikalische Modellierungskonzepte. Springer, Berlin; Auflage: 1 (Januar 1997), ISBN: 3540619275.

- HOOGENDOORN, S.; BOVY, P. [2001]: State-of-the-art of Vehicular Traffic Flow Modelling, Special Issue on Road Traffic Modelling and Control of the Journal of Systems and Control Engineering. Volume 215, Number 4 / 2001, S. 283-303, ISSN 0959-6518.
- HOOPS, M. [2003]: Methodik zur Qualitätssicherung von aggregierten Daten eines Messsystems im Betrieb von verkehrstechnischen Anlagen. Dissertation am Lehrstuhl für Verkehrstechnik der TU München, 2003.
- HOUCK, C.; JOINES, J.; KAY, M. [1995]: A Genetic Algorithm for Function Optimization: A Matlab Implementation. NCSU-IE TR 95-09, 1995.
- HUBER, W. [2001]: Fahrzeuggenerierte Daten zur Gewinnung von Verkehrsinformationen. Dissertation am Fachgebiet Verkehrstechnik und Verkehrsplanung der TU München, 2001.
- HUISKEN, G. [2006]: Inter-Urban Short Term Traffic Congestion Prediction. Dissertation an der Universität von Twente, Niederlande, 2006.
- JAIN, R.; MAZUMDAR, J. [2003]: A genetic algorithm based nearest neighbor classification to breast cancer diagnosis. Australas Phys Eng Sci Med. 2003 Mar;26(1):6-11.
- JANKO, J. [1994]: Probleme der Reisezeitprognose in einem Leitsystem für den Straßenverkehr. PhD Thesis at the TU Berlin, February 1994.
- JÖNSSON, P.; WOHLIN, C. [2004]: An Evaluation of k-Nearest Neighbour Imputation Using Likert Data. Proceedings of the 10th International Symposium on Software Metrics (METRICS'04).
- JUNG, B. [2008]: Syntaktische und Statistische Mustererkennung. VO 502 258, Vorlesungsfolien SS2008 2. Teil (13.3.2008).
- KAKDE, H. [2005]: Range Searching using Kd Tree. 25-08-2005
- KAPARIAS, I.; BELL, M.G.H.; BOGENBERGER, K.; CHEN, Y. [2007]: An approach to time-dependence and reliability in dynamic route guidance, Proceedings of the 86th Annual Meeting of the Transportation Research Board., 2007.
- KATES, R., BOBINGER R., STOICA, I.; MÜCK, J. [2005]: Fusion von Verkehrsprognosen für die Praxisanwendung, Straßenverkehrstechnik, 49 (6), S. 296-302.
- KEMPER, C. [2006]: Dynamische Simulation des Verkehrsablaufs unter Verwendung statischer Verflechtungsmatrizen. Dissertation an der Fakultät für Bauingenieurwesen und Geodäsie der Universität Hannover. 2006.
- KERNER, B. S.; REHBORN, H.; ALEKSIC, M., HAUG, A. [2001]: Methods for tracing & forecasting congested traffic patterns. TEC 2001, S. 282-287.
- KERNER, B. S. [2004]: The Physics of Traffic, Springer, Heidelberg 2004, ISBN: 3-540-20716-3.
- KIMBER, R.; HOLLIS, E. [1979]: Traffic Queues and Delays at Road Junctions. TRRL Laboratory Report, 909, U.K.
- KIRSCHFINK, H.; CHADENAS, C. [1998]: Traffic Situation Prediction Applying Pattern Matching and Fuzzy Classification, Aachen 1998.
- KLEINHANS, M. [2002]: Klassifikation von Verkehrsdaten mit Clusterverfahren. Diplomarbeit am Mathematischen Institut der Universität Köln.
- KÜSTERS, U. [2001]: Data Mining Methoden - Einordnung und Überblick. Skriptum zum Seminar "Data Mining" an der Wirtschaftswissenschaftlichen Fakultät Ingolstadt der Katholischen Universität Eichstätt.

- LEONHARDT, A. [2006]: Verkehrslageschätzung in Straßennetzen durch Kopplung streckenbezogener und querschnittsbezogener Verkehrsdaten, Hochschultagung Straßen- und Verkehrswesen, September 2006 in Hohenwart.
- LEONHARDT, A. [2007]: Güte und Aussagekraft von Reisezeiten Taxipositionsmeldungen. Internes Arbeitspapier, Lehrstuhl für Verkehrstechnik, TU München.
- LERNER, G. [2005]: Extended Floating Car Data – Mobile Incident Detection for Safety & Efficiency Applications enabled by Car 2 X Communication.
- LINAUER, M.; DIN, K.; ASAMER, J.; NOWOTNY, B.; NORMAN, N.; HEININGER, M.; HEININGER, D.; PAUKERL, A.; TUREK, T.; STEINER, P. [2006]: FUSION - Floating Car Daten und Sensordaten intelligent fusionieren. Endbericht im Rahmen von I2 - Intelligente Infrastruktur. bmvit.
- LISTL, G. [2003]: Anwendung neuer Technologien zur Erfassung des Verkehrsablaufs. Dissertation an der Universität Kassel, Mai 2003.
- LIU, H.; BAN, X.; RAN, B.; MIRCHANDI, P. [2001]: An Analytical Dynamic Traffic Assignment Model with Probabilistic Travel Times and Perceptions, UCI-ITS-WP-01-14.
- LÖWE, S. [2004]: Rechnergestützte Betriebsleitsysteme. Seminar on Car-2-Car Communication (C2CC), Lehrstuhl Systeme, BTU Cottbus.
- LOGI, F. [2002]: Effiziente Routensuche in einer verteilten Umgebung. Forschungsbericht im Rahmen des BMBF Projektes DIRECT am Fachgebiet Verkehrstechnik und Verkehrsplanung der TU München, 2002.
- LOGI, F.; ULLRICH, M. [2001]: The Dynamic Network Model DINO: Documentation, Version 1.5 (03.07.01), working paper, Fachgebiet Verkehrstechnik und Verkehrsplanung, TU München.
- LORKOWSKI, S.; BROCKFELD, L.; MIETH, P.; PASSFELD, B.; THIESSENHUSEN, K.-U.; SCHÄFER, R.-P. [2003]: Erste Mobilitätsdienste auf Basis von "Floating Car Data". In: RWTH Aachen [Hrsg.]: Tagungsband zum 4. Aachener Kolloquium "Mobilität und Stadt", Stadt Region Land, 75, S. 93 - 100, 4. Aachener Kolloquium "Mobilität und Stadt" AMUS, Aachen (Deutschland), 2003-07-31 - 2003-08-01 (2003).
- MAIER, F.; BUSCH, F.; ROTH, C. [2007]: Verfahren zur Wiedererkennung von Fahrzeugen anhand der Signaturen über Induktivschleifen. Straßenverkehrstechnik 02/2008.
- MAIER, F.; BRAUN, R.; BUSCH, F.; MATHIAS, P. [2008]: Pattern-based short-term prediction of urban congestion propagation and automatic response. Traffic Engineering & Control, TEC Magazine 06/2008.
- MESCHENDÖRFER, J. [2004]: Fahrzeugklassifizierung und Fahrzeugwiedererkennung anhand von Mikrodaten lokaler Detektoren. Diplomarbeit am Lehrstuhl für Verkehrstechnik der TU München.
- MINGZHE, L., RUILI, W., JIANGSHENG, W., KEMP, R. [2005]: A genetic-algorithm-based neural network approach for short-term traffic flow forecasting. Advances in neural networks - ISSN 2005 (Chongqing, 30 May - 1 June 2005) vol. 3498, ISBN 3-540-25912-0.
- MOORE, A. [2001]: Instance-based learning (a.k.a. memory-based) (a.k.a. nonparametric regression) (a.k.a. casebased) (a.k.a. kernel-based). Tutorium, School of Computer Science Carnegie Mellon University.
- NAKATA, T.; TAKEUCHI, J. [2004]: Mining Traffic Data from Probe-Car System for Travel Time Prediction. KDD'04, August 22-25, 2004, Seattle, Washington, USA.
- NANTHAWICHIT, C.; NAKATSUJI, T.; SUZUKI, H. [2003]: Application of Probe Vehicle data for Real-Time Traffic State Estimation and Short-term Travel Time Prediction on a Freeway. TRB 2003 annual Meeting, Washington D.C., USA.

- NATH, A. K.; RAHMAN, S. M.; SALAH, A. [2005]: An Enhancement of k-Nearest Neighbor Classification Using Genetic Algorithm. Working Paper at Department of Computer Science, North Dakota State University.
- NIEMANN, H. [2003]: Klassifikation von Mustern (2. überarbeitete und erweiterte Auflage im Internet www5.informatik.uni-erlangen.de/our-team/niemann-heinrich/books.html, Stand 14.06.2007). Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg.
- NOLAND, R.; SMALL, K.; KOSKENOJA, X. [1998]: Simulating Traveltime Reliability. *Regional Science and Urban Economics* 28 (1998) 535-564.
- KAMARIANAKIS, Y.; PRASTACOS, P. [2002]: Space-time modeling of traffic flow. ERSA conference papers ersa02p141, European Regional Science Association, revised.
- OFFERMANN, F. [2003]: Ein Neuro-Fuzzy-Modell zur Reisegeschwindigkeitsabschätzung auf Richtungsfahrbahnen basierend auf einer Fusion lokaler und fahrzeuggenerierter Verkehrsdaten. Dissertation an der RWTH Aachen, 2003.
- OSWALD, K.; SCHERER, W.; SMITH, B. [2000]: Traffic flow forecasting using approximate nearest neighbor nonparametric regression. Research Report No. UVACTS-15-13-7, December 2001. Center for Transportation Studies at the University of Virginia.
- PFANNERSTILL, E. [1983]: Ein Verfahren zur Merkmalsextraktion aus Linienmustern zur Zuordnung von Fahrzeugkollektiven. Dissertation an der RWTH Aachen, 1983.
- PRIEMER, C.; FRIEDRICH, B. [2008]: Optimierung von modellierten Warteprozessen im Rahmen adaptiver Netzsteuerungen durch C2I-Daten. Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.), HEUREKA '08, Optimierung in Verkehr und Transport - Tagungsband, FGSV Verlag, Köln, ISBN 978-3-939715-48-1.
- RAU-BEDROW, H. [2002]: Value at Risk, Normalverteilungshypothese und Extremwertverhalten. *Finanz Betrieb, Zeitschrift für Unternehmensfinanzierung und Finanzmanagement*, 3. Jahrgang, Oktober 2002, S. 603 – 607.
- RICE, J.; VAN ZWET, E. [2001]: A simple and effective method for predicting travel times on free-ways. *Intelligent Transportation Systems*, 2001. Proceedings. 2001 IEEE Volume, Issue, 2001 Page(s):227 – 232.
- ROBINSON, S. [2005]: The development and application of an urban link travel time model using data derived from inductive loop detectors. PhD Thesis, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom, August 2005.
- ROBINSON, S.; POLAK, J. W. [2005]: Modelling Urban Link Travel Time with Inductive Loop Detector data using the k-NN method, 84th Transportation Research Board Annual Meeting, Jan. 9 - 13, 2005, Washington, D.C. and publication in the *Transportation Research Record*.
- ROBINSON, S.; POLAK, J. W. [2006]: ANPR travel time data filtering using the Overtaking Rule at low flows. Proceedings of the ITS World Congress 2006, London.
- ROBINSON, S.; POLAK, J.W. [2007]: Characterising the components of Urban Link Travel Time Variability Using the k-NN (k-Nearest Neighbors) method, presented at the 86th Annual Meeting of the Transportation Research Board (2007).
- RITCHIE, S.; SUN, C. [2001]: Section Related Measures of Traffic System Performance: Final Report. Institute of Transportation Studies, University of California, Berkeley. California PATH Research Report, UCB-ITS-PRR-98-33.

- SACHSE, T. [2000]: Alternativroutensteuerung in Autobahnnetzen auf der Basis einer erweiterten Analyse des Verkehrsablaufs. Dissertation an der TU München, 2000.
- SARLE, W. S. [1994]: Neural Networks and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, April 1994.
- SCHÄFER, R. [2005]: Real-time Floating Car Data – A data source generating TMC messages. TMC Workshop, Beijing, 5. April 2005.
- SCHIMANDL, F.; LEONHARDT, A. [2008]: NOSTRADAMUS - Real Time Traffic Data Processing for Large Probe Vehicle Data Sets. Use Case: Taxi FCD Server Guangzhou, internes Arbeitspapier, Lehrstuhl für Verkehrstechnik, Technische Universität München, 2008.
- SCHMIDT, D. [2007]: Verkehrslagererkennung in Dmotion, Datenfusion mit DINO. Workshop Verkehrslageanalyse mit Hilfe von Simulationsmodellen - erreichter Stand und Perspektiven" der ASIM-Fachgruppe 4.5.8, Simulation von Verkehrssystemen, 08.02.07, Weimar.
- SCHNÖRR, C. [2005]: Numerische Einkopplung von Floating-Car-Daten in Makroskopische Verkehrsmodelle. Straßenverkehrstechnik 12/2005.
- SCHWERDTFEGER, T. [1986]: Makroskopisches Simulationsmodell für Schnellstraßennetze mit Berücksichtigung von Einzelfahrzeugen (DYNEMO). Dissertation and der TH Karlsruhe, 1986.
- SHEFFI, Y. [1985]: Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, New Jersey.
- SIEKMANN, S. [1999]: Fusion von Expertenwissen und Daten mit Neuro-Fuzzy-Methoden zur Prognose von Finanzzeitreihen. Dissertation an der Universität Marburg, 1999.
- SMITH, B.; DEMETSKY, M. [1997]: Traffic Flow Forecasting: Comparison of Modeling Approaches. Journal of Transportation Engineering 123(4), 261–266.
- SPANGLER, M. [2005]: Reisezeitermittlung in Abhängigkeit von Verkehrsbelastungen im Münchener Norden am Beispiel der Ingolstädter Straße. Endbericht im Auftrag der BMW AG, München, Januar 2005.
- SPANGLER, M. [2008]: Online Erfassung von Reisezeiten mittels Kennzeichenerfassung in München. Internes Arbeitspapier, Lehrstuhl für Verkehrstechnik, Technische Universität München, Juni 2008.
- STEINER, A.; SICK, B. [2007]: New Methods for Travel Time Estimation on Freeway Sections. Swiss Transport Research Conference, 12-14 September 2007, Ascona, Switzerland (Hrsg.).
- STEINER, A.; SICK, B. [2008]: A New Method for Travel Time Estimation on Long Freeway Sections. European Journal of Transport and Infrastructure Research, accepted for Publication.
- SWIDER, J.; HASCHKE, B.; RUDION, K.; HEYDE, C. [2006]: Verteilungsdichteprognose von Netzlast und Preisen, AP 1.4 im Projekt NetMod.
- TANTIYANUGULCHAI, S. AND BERTINI, R. [2003]: Arterial Performance Measurement Using Transit Buses as Probe Vehicles. Portland State University Department of Civil & Environmental Engineering. 0-7803-8125-4/03/\$17.00 © 2003 IEEE.
- TAO, Y.; YANG, F. QIU, Z.; RAN, B. [2006]: Travel Time Prediction in the Presence of Traffic Incidents Using Different Types of Neural Networks. TRB 2006 Annual Meeting CD-ROM.
- TATINENI, M. BOYCE, D.; MIRCHANDANI, P. [1997]: Experiments to Compare Deterministic and Stochastic Network Traffic Loading Models. Transportation Research Record, 1607, pp. 16-23.
- TLS [2006]: Technische Lieferbedingungen für Streckenstationen, Ausgabe 2006.

- TSEKERIS, T.; STATHOPOULOS, A. [2006]: Real-Time traffic volatility forecasting in urban arterial Networks. TRB 2006 Annual Meeting.
- TUROCHY, R. E.; PIERCE B. D. [2004]: Relating Short-Term Traffic Forecasting to Current System State Using Nonparametric Regression, 2004 IEEE Intelligent Transportation systems Conference Washington, D.C., USA, October 34, 2004.
- VAN LINT, H. [2004]: Reliable Travel Time Prediction for Freeways, PhD Thesis at the TU Delft, The Netherlands, 2004.
- VAN ZUYLEN, H.J. [1981]: Some improvements in the estimation of an OD matrix from traffic counts. Proceedings of the 8th International Symposium on Transportation and Traffic Theory, University of Toronto Press, Toronto, Canada, pp. 183-190.
- VON DER RUHREN, S. [2000]: Verfahren zur Klassifizierung von Ganglinien, Arbeitspapier F4, Institut für Stadtbauwesen und Stadtverkehr, RWTH Aachen, Oktober 2000.
- VON DER RUHREN, S. [2006]: Kurzfristprognosen von Verkehrszuständen auf Basis von Verfahren der Mustererkennung und von dynamischen Routensuch- und Umlegungsverfahren", Dissertation am Institut für Stadtbauwesen und Stadtverkehr, RWTH Aachen, Juli 2006.
- VORTISCH, P. [2005]: Modellunterstützte Messwertpropagierung zur Verkehrslageschätzung in Stadtstraßennetzen. PhD Thesis at the TH Karlsruhe (2005).
- VORTISCH, P.; MENNENI, S.; SUN, C. [2008]: Kalibrierung von Fahrverhaltensparametern in der mikroskopischen Verkehrsflusssimulation mit Hilfe evolutionärer Algorithmen. Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.), HEUREKA '08, Optimierung in Verkehr und Transport - Tagungsband, FGSV Verlag, Köln, ISBN 978-3-939715-48-1.
- WEICKER, K. [2007]: Evolutionäre Algorithmen. 2. Auflage, Teubner Verlag. ISBN 978-3-8351-0519-4.
- WEINZERL [2006]: B.L.I.D.S – Bluetooth-based Traffic Data Collection System, internal presentation, 2006.
- WILD, D. [1996]: Die Prognose von Verkehrsstärken anhand klassifizierter Ganglinien; Dissertation an der Universität Karlsruhe, 1996; Shaker-Verlag, Aachen, ISBN 3-8265-1478-5
- WU D.; ZHU, T.; WANG, G.; DU, Y. [2006]: A quick Map-Matching Algorithm for large-scale Float Car Data. Proceedings of the ITS World Congress 2006, London.
- YANG, Z.; BAO, L. ZHU, G. [2000]: An Urban Freeway Travel Time Prediction Model based on Fuzzy Linear Regression.
- YOO, B.; KANG, S.; PARK, C. [2005]: Travel Time Estimation using Mobile Data. Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5 pp. 1533-1547, 2005.
- YOU, J.; KIM, T. [2000]: Development and Evaluation of a hybrid travel time estimation model. Transportation Research C (8), S. 231–256.
- ZHOU, C.; NELSON, P. [2002]: Predicting Traffic Congestion using Recurrent Neural Networks. Proceedings of 9th World Congress on Intelligent Transport Systems, Chicago, Oktober 2002.
- ZILIASKOPOULOS, A.; MAHMASSANI, H. [1996]: A Note on least Time Path Computation Considering Delays and Prohibitions for Intersection Movements. Transportation Research B, Vol. 30. No. 5, pp. 359-367, 1996.
- ZITZELSBERGER, M. [2006]: Mustererkennung unter Verwendung der Stützvektormethode, Diplomarbeit an der Fachhochschule Regensburg, Fachbereich Informatik/Mathematik, März 2006.

Abbildungsverzeichnis

Abb. 1.1	Methodischer Aufbau der Arbeit und des Verfahrens	4
Abb. 2.1	Darstellung des Belegungsgrades (Median über 15 Minuten) je Detektor über die Zeit. Graz, Mittwoch 02.05.2007	9
Abb. 2.2	Lokaler Belegungsgrad über Zeit (21.06.2007. Det. ID 622). Darstellung von Rohdaten, aggregierten Daten und exponentiell geglätteten Daten für verschiedene α	12
Abb. 2.3	Fahrzeugwiedererkennung (prinzipielle Systemanordnung)	15
Abb. 2.4	Münchner Kennzeichenkamera-Messstellennetz [SPANGLER 2008]	17
Abb. 2.5	Darstellung der potentiell nutzbaren Taxifahrten vom 13.03.2007. Links: München. Rechts: Verbindung der nutzbaren Taxipositionen im Grazer Hauptstraßennetz.	20
Abb. 2.6	Darstellung einer Taxifahrt.....	20
Abb. 2.7	Darstellung des Verfahrens zur Ermittlung von Reisezeiten aus Fahrzeugpositionsmeldungen	21
Abb. 2.8	Illustration der Berechnung der wahrscheinlichsten Route	22
Abb. 2.9	Testgebiet Graz	25
Abb. 2.10	Testgebiet München.....	25
Abb. 2.11	Darstellung der Filterwirkung bei Anwendung auf mittels Kennzeichenerfassung ermittelten Reisezeiten (5.-6.09.2007/Petuelring – Ingolstädter Straße/München)	26
Abb. 2.12	Reisezeiten auf der A9.....	26
Abb. 2.13	Reisezeiten zwischen Petuelring und Freimann (B2R).	26
Abb. 2.14	Zusammenhang zwischen Routenlänge, durchschnittlicher Anzahl von Beobachtungen pro Tag und möglicher relativer Abweichung von zwei aufeinander folgend erhobenen Einzelfahrzeugreisezeiten.....	27
Abb. 2.15	Belegungsgrad-Verkehrsstärke Diagramme (basierend auf 15-Minutenwerten).....	29
Abb. 2.16	Schätzung und Prognose von Reisezeiten aus lokalen Verkehrsstärkemessungen (München, Reisezeit (rot) und Verkehrsstärke (blau) eines LSA-Detektors auf der Route "INGOLSTÄDTER")	29
Abb. 2.17	Zeitabhängigkeit der Reisezeit (links, INGOLSTÄDTER) und des Belegungsgrades (rechts, ERZHERZOG).....	30
Abb. 2.18	Darstellung des Belegungsgrades als Funktion des Belegungsgrades 15 Minuten früher am gleichen Detektor.....	31
Abb. 2.19	Darstellung des Belegungsgrades eines Detektors B als Funktion des Belegungsgrades eines stromaufwärts gelegenen Detektors A.....	31
Abb. 3.1	Ablauf der Mustererkennung (in Anlehnung an NIEMANN [2003]).....	34
Abb. 3.2	Klassifizierung und Segmentierung der Ganglinienbasis [WILD 1996].....	36
Abb. 3.3	Ablauf der Regressionsmodellbildung.....	38

Abb. 3.4	Zur Prognose verwendete Messstellen (links) und deren Repräsentation als Baum (rechts) [KAMARIANAKIS UND PRASTACOS 2002].....	40
Abb. 3.5	Mustererkennung mit künstlichen Neuronalen Netzen	40
Abb. 3.6	Selbstlernendes Expertensystem.....	42
Abb. 3.7	Gelernter Zusammenhang zwischen v_{lokal} , q und v_{Route} [LINAUER ET AL. 2006].....	43
Abb. 3.8	Instanzbasiertes Lernen	43
Abb. 3.9	Verkehrsstärke und transformierte Ganglinie [WILD 1996]	44
Abb. 3.10	Prinzipskizze des Prognoseverfahrens von FÖRSTER [2008]	45
Abb. 3.11	Optimierung von Verfahren zur Mustererkennung	47
Abb. 4.1	Schätzung der Verkehrsstärke aus der Geschwindigkeit an einem lokalen Messquerschnitt mit Instanzbasiertem Lernen.	52
Abb. 4.2	Schätzung der Belegung aus der Tageszeit bei Vorliegen verschiedener Tagesklassen (Datenquelle: Graz, März/April 2007)	53
Abb. 4.3	Schätzung von $y' = f(x)$ mittels 3 Nächsten Nachbarn (3NN)	53
Abb. 4.4	Schematische Darstellung des Mustererkennungsprozesses	54
Abb. 4.5	Darstellung der Merkmalsgewichtung als Funktion von Ort und Zeit	58
Abb. 4.6	Gewichtsfunktionen g für verschiedene Parameter ds_{max} und dt_{max} (dt ist der Interpretierbarkeit wegen in Minuten dargestellt).....	58
Abb. 4.7	Darstellung der Repräsentanten für verschiedene Tagesklassen für zwei Detektoren in Graz (Zeitraum: März 2007 – Mai 2008)	60
Abb. 4.8	Darstellung der Distanzen D_k für sonntags 00:00-00:15 Uhr (links) und montags 07:00-07:30 Uhr (rechts), Datenbasis: Belegungsgrade "EXERZIERPLATZ"	61
Abb. 4.9	Ermittlung der Anzahl der gültigen Merkmale des Verkehrsmusters	62
Abb. 4.10	Darstellung verschiedener Kernfunktionsformen K	63
Abb. 4.11	Reisegeschwindigkeit auf einer Route als Funktion des Belegungsgrades eines Detektors	65
Abb. 4.12	Darstellung unterschiedlich zuverlässiger Schätzungen	67
Abb. 4.13	Schätzung von v aus q und umgekehrt für $k = 51$, Daten der A9 München – Nürnberg km 518,530 (AS Garching-Nord), 25. Februar 2006	70
Abb. 4.14	Beispiel zur Abschätzung der Überschreitungswahrscheinlichkeit für die Reisezeit für $k = 13$ Nachbarn.....	72
Abb. 4.15	Gesamtübersicht über das Verfahren	73
Abb. 4.16	Einbettung des Prognosemoduls in den Datenfluss und das Dateisystem.....	74
Abb. 5.1	Vorgehen bei Sensitivitätsanalyse, Parameteroptimierung und Bewertung des Verfahrens.....	77
Abb. 5.2	Darstellung der Abhängigkeit des Erwartungswertes von den Parametern ds_{max} und dt_{max} für die Schätzung des Belegungsgrades (links) und der Reisezeit (rechts)	81

Abb. 5.3	Abhängigkeit von G_V vom Prognosehorizont t_{prog}	82
Abb. 5.4	Darstellung des Einflusses von k auf den Erwartungswert für vier Zeitpunkte t_1 bis t_4 für die Belegungsgradprognose (links) und die Reisezeitprognose (rechts)	83
Abb. 5.5	Darstellung des Einflusses des Parameters K_w bei der Gewichtung der Instanzen als Funktion der Distanz der Instanz zum aktuellen Muster (links) und als Funktion des Rangs des Nachbarn (rechts).....	84
Abb. 5.6	Darstellung der Streuung des Erwartungswertes des Belegungsgrades für verschiedene Prognosehorizonte bei Variation der Parameter k (links) und K_w (rechts)	87
Abb. 5.7	Darstellung der Streuung des Erwartungswertes der Reisezeit für verschiedene Prognosehorizonte bei Variation der Parameter k (links) und K_w (rechts).....	88
Abb. 5.8	Ablauf der Parameteroptimierung mit einem Genetischen Algorithmus.....	89
Abb. 5.9	Visualisierung des Zielfunktionswertes MQF bei der Optimierung der Belegungsgradprognose (links) und der Reisezeitprognose (rechts) mit GA	92
Abb. 5.10	MLF zur Abbildung der Zusammenhänge des Belegungsgrades an einem Detektor (y) durch die Belegungsgrade der vier letzten Zeitintervalle zweier stromaufwärts gelegener Detektoren (x_1 und x_2)	97
Abb. 5.11	Performanz des MLF als Funktion der Trainingsepochen für die 15-Minuten und 45-Minuten-Prognose des Belegungsgrades EXERZIERPLATZ (links) und der Reisezeit OPERNRING (rechts)	98
Abb. 5.12	Darstellung von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 30 Minuten (rechts), EXERZIERPLATZ.....	99
Abb. 5.13	Vergleich von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 45 Minuten (rechts), EXERZIERPLATZ	100
Abb. 5.14	MPF (li.) und MQF (rechts) zwischen Prognose und Referenzwert, EXERZIERPLATZ	101
Abb. 5.15	Darstellung von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 30 Minuten (rechts), ERZHERZOG.....	102
Abb. 5.16	Vergleich von prognostiziertem Belegungsgrad und Referenzwert für einen Horizont von 15 Minuten (links) und 45 Minuten (rechts), ERZHERZOG	103
Abb. 5.17	MPF (li.) und MQF (re.) zwischen Prognose und Referenz, ERZHERZOG.....	104
Abb. 5.18	Einfluss der Trainingsdatenbasis auf das Schätzergebnis	105
Abb. 5.19	Verlauf der prognostizierten Reisezeiten und zugehörige Taxireisezeiten für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), OPERNRING	107
Abb. 5.20	Vergleich von prognostizierter Reisezeit und Taxireisezeiten für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), OPERNRING.....	108
Abb. 5.21	Verlauf der prognostizierten Reisezeiten und zugehörige Taxireisezeiten für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), OPERNRING	109
Abb. 5.22	Vergleich von prognostizierter Reisezeit und Taxireisezeiten für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), KEPPLER-WEST	110

Abb. 5.23	Verlauf der prognostizierten Reisezeiten und zugehörige Reisezeiten aus Fahrzeugwiedererkennung für Prognosehorizonte 0 Minuten (oben) und 15 Minuten (unten), INGOLSTÄDTER.....	111
Abb. 5.24	Vergleich von prognostizierter Reisezeit und Reisezeiten aus Fahrzeugwiedererkennung für einen Horizont von 0 Minuten (links) und 30 Minuten (rechts), INGOLSTÄDTER.....	112
Abb. 5.25	Geschätzte Reisezeit unter Verwendung von Heuristiken (OPERNRING).....	115
Abb. 5.26	Vergleich von Reisezeitschätzung und Taxireisezeiten ($t_{\text{Prog}} = 0$), Reaktion der Zuverlässigkeitsindikatoren (OPERNRING).....	116
Abb. 5.27:	Darstellung der Überschreitungswahrscheinlichkeit der Reisezeit über die Zeit für verschiedene Prognosehorizonte und kalenderbasiert auf dem "OPERNRING".....	118
Abb. 6.1	Verkehrsqualitätskarten für Graz.....	123

Tabellenverzeichnis

Tab. 2.1	Dauer des Map Matching Vorgangs für verschiedene Matching Verfahren	21
Tab. 2.2	Untersuchungsgebiet Graz, Untersuchungsgegenstände der einzelnen Lokationen	24
Tab. 2.3	Untersuchungsgebiet München, Untersuchungsgegenstände der einzelnen Lokationen	25
Tab. 3.1	Modellklassen und zugehörige Entwurfparameter	47
Tab. 5.1	Grundeinstellung der Parameter	81
Tab. 5.2	Absolute Streuungen der Prognose je Parameter und Situation	86
Tab. 5.3	Relative Streuungen der Prognose je Parameter und Situation	86
Tab. 5.4:	Optimierte Parameter	94
Tab. 5.5	Korrelation zwischen Prognose und Referenz, Belegungsgrad, EXERZIERPLATZ	101
Tab. 5.6	MAF zwischen Prognosewert und Referenz, Belegungsgrad, EXERZIERPLATZ....	102
Tab. 5.7	Korrelation zwischen Prognosewert und Referenz, Belegungsgrad, ERZHERZOG	104
Tab. 5.8	MAF zwischen Prognosewert und Referenz, Belegungsgrad, ERZHERZOG.....	104
Tab. 5.9	Korrelation zwischen Prognosewert und Taxireisezeit, OPERNRING	108
Tab. 5.10	Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Taxireisezeit, OPERNRING (tolerierte Abweichung: 60 Sekunden).....	109
Tab. 5.11	Korrelation zwischen Prognosewert und Taxireisezeit, KEPLER-WEST"	110
Tab. 5.12	Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Taxireisezeit, KEPLER-WEST (tolerierte Abweichung: 60 Sekunden)	110
Tab. 5.13	Korrelation zwischen Prognosewert und Reisezeit auf Basis von Fahrzeugwiedererkennung, INGOLSTÄDTER	112
Tab. 5.14	Mittlere absolute Abweichung [Sekunden] zwischen Prognosewert und Reisezeit auf Basis von Fahrzeugwiedererkennung, INGOLSTÄDTER (tolerierte Abweichung: 60 Sekunden)	112

Anhang 1

Verfahren zur Kürzestwegsuche

Bei Routensuchverfahren (Kürzestwegsuchverfahren, *engl.: Shortest Path Algorithm*) kann grundsätzlich zwischen uninformatierter und informierter Suche unterschieden werden. Der vermutlich bekannteste Vertreter der uninformatierten Suche ist der Dijkstra Algorithmus (beschrieben von DIJKSTRA [1959]). Dijkstra's Algorithmus findet jeden existierenden kürzesten Weg in einem gerichteten Graphen mit nicht negativen Kantengewichten (was in einem Straßennetz gegeben ist). Grundsätzlich sucht der Algorithmus dabei ausgehend von der Quelle in alle Richtungen mit dem Ziel, die Kosten (Routenlänge) möglichst niedrig zu halten. Die Suche wird abgebrochen, wenn das Ziel erreicht ist. Bei der Suche sind Kanten, die in Richtung des Ziels führen, und Kanten, die vom Ziel weg führen, gleich berechtigt (daher *uninformierte* Suche). Der Dijkstra Algorithmus sucht mehr oder weniger kreisförmig um den Startpunkt herum.

Im Gegensatz zur uninformatierten Suche nutzt die informierte Suche Informationen über die Lage des Ziels im Vergleich zur Quelle. Dabei werden bei der Suche diejenigen Kanten bevorzugt, die in Richtung der Quelle liegen. Der wahrscheinlich am weitesten verbreitete Vertreter der informierten Suchverfahren ist der A* ("A Stern") - Algorithmus. Er wird z.B. in aktuellen Navigationssystemen genutzt und ist Gegenstand ständiger Forschung und Erweiterungen (z.B. CHEN ET AL. [2005], KAPARIAS ET AL. [2007]).

Beim A* Algorithmus wird neben den Kosten, einen Punkt p zu erreichen, eine Schätzung der Distanz zum Erreichen des Ziels von p aus angesetzt. Dadurch werden nicht nur die Kosten zum Erreichen des Punktes p betrachtet, sondern mit der potentiellen Distanz auch Abschätzung der Kosten, um von dort zum Ziel zu gelangen.

Der kritische Punkt ist die Schätzung der Kosten von einem Punkt p zum Ziel. Auf der einen Seite dürfen die wahren Kosten nicht überschätzt werden, da ansonsten nicht garantiert werden kann, dass der kürzeste Weg gefunden wird. Auf der anderen Seite ist der Algorithmus umso effizienter, je näher die Schätzung an der wahren Distanz liegt (*wird die Distanzschätzung zu 0 gesetzt, mutiert der A* Algorithmus zum Dijkstra Algorithmus*). Im praktischen Gebrauch hat sich die euklidische Distanz als Schätzung der wahren Distanz im Straßennetz als sinnvoll erwiesen.

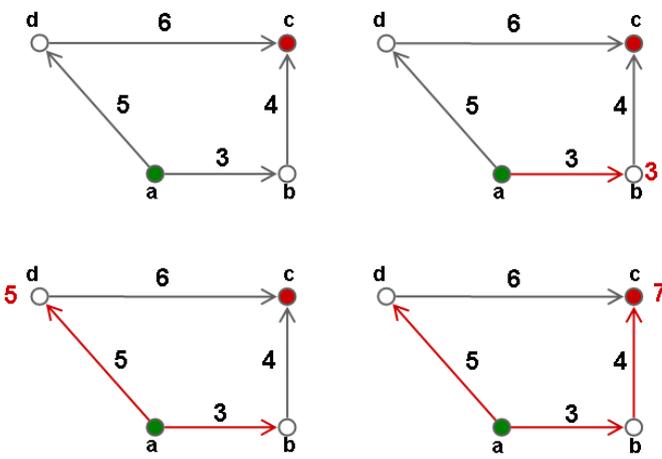
Berechnungsbeispiele

In den unten stehenden Abbildungen werden die Funktionsweise des Dijkstra Algorithmus bzw. des A* Algorithmus anhand eines einfachen Beispielnetzes erläutert. Gesucht wird jeweils die kürzeste Route zwischen den Knoten (a) und (c).

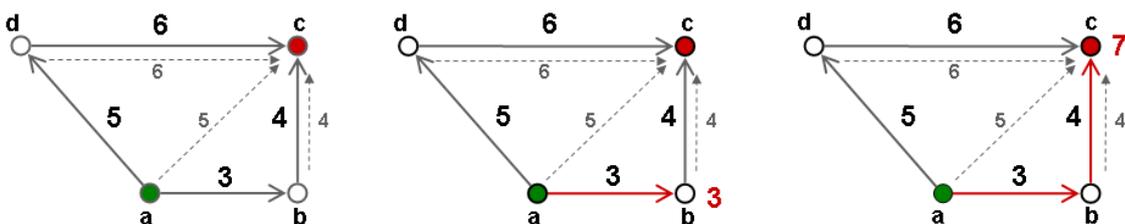
Unter Nutzung des Dijkstra Algorithmus wird zunächst Kante (ab) genutzt, da die Kosten am niedrigsten sind. Als nächstes wird Kante (ad) benutzt, da die Kosten mit 5 niedriger sind, als direkt die Kante (bc) zu nutzen (Kosten = 3+4 = 7). Im dritten Schritt wird die Kante (bc) genutzt und die Suche ist abgeschlossen, da das Ziel erreicht ist.

Mit dem A* Algorithmus wird zunächst ebenfalls Kante (ab) benutzt. Im zweiten Schritt werden die Kosten der Nutzung von (ad) $5 + 6 = 11$ geschätzt, wobei 6 gleich der euklidischen Distanz zwischen (d) und (c) ist. Die Kosten der Nutzung von (bc) ergeben sich zu $3 + 4 + 0 = 7$, wobei 0 die geschätzte Entfernung von (c) zum Ziel ist. Konsequenterweise wird direkt die Kante (bc) genutzt und der Algorithmus terminiert.

In diesem einfachen Beispiel benötigt der A* Algorithmus zwei und der Dijkstra Algorithmus drei Berechnungsschritte. Der Vorteil für den A* Algorithmus wird größer, je weiter Quelle und Ziel auseinander liegen.



Darstellung des Dijkstra Algorithmus



Darstellung des A* Algorithmus

Anhang 2

Ergebnisse der Parameteroptimierung

Exerzierplatz

Generation	G_V	k_e	dt_{\max}	ds_{\max}	k	K_W	MQF
1	9.82	1.75	4.57	2.42	172	1.4	0.33

Ergebnisse der Parameteroptimierung, Szenario "EXERZIERPLATZ", $t_{\text{Prog}} = 15$ Minuten

Generation	G_V	k_e	dt_{\max}	ds_{\max}	k	K_W	MQF
1	9.81	1.75	4.57	2.42	172	1.4	0.46
4	10	1.51	10	2.47	101	4.9713	0.45

Ergebnisse der Parameteroptimierung, Szenario "EXERZIERPLATZ", $t_{\text{Prog}} = 30$ Minuten

Generation	G_V	k_e	dt_{\max}	ds_{\max}	k	K_W	MQF
1	5.48	1.58	8.21	1.46	88	6.22	0.72
3	7.48	1.74	6.61	2.74	25	9.53	0.58
8	5.11	1.74	6.61	2.74	25	9.53	0.57

Ergebnisse der Parameteroptimierung, Szenario "EXERZIERPLATZ", $t_{\text{Prog}} = 45$ Minuten

ERZHERZOG

Generation	G_V	k_e	dt_{\max}	ds_{\max}	k	K_W	MQF
1	7.71	9.83	3.36	1.32	240	2.53	0.89
3	5.33	2.61	5.30	2.24	57	3.26	0.82
4	10	1	1	1	101	10	0.60
6	10	1	5.30	2.43	56	6.28	0.46

Ergebnisse der Parameteroptimierung, Szenario "ERZHERZOG", $t_{\text{Prog}} = 15$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	7.71	9.83	3.36	1.32	240	2.53	1.13
3	10	7.54	1	3.82	301	10	1.10
4	10	1	1	1	101	10	0.88
6	10	1	2.57	2.15	222	9.31	0.82
9	8.15	1	3.52	3.08	48	5.70	0.76

Ergebnisse der Parameteroptimierung, Szenario "ERZHERZOG", $t_{prog} = 30$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	7.71	9.83	3.36	1.32	240	2.53	1.27
3	10	7.54	1	3.82	301	10	1.22
4	10	1	1	1	101	10	1.05
6	0.28	1	1	1	1	0.0001	0.95

Ergebnisse der Parameteroptimierung, Szenario "ERZHERZOG", $t_{prog} = 45$ Minuten

OPERNRING

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	8.74	1.39	2.36	3.17	71	3.45	1733.88
4	9.02	8.41	2.30	2.44	117	4.31	1733.02
9	9.02	8.40	1.30	3.70	32	0.19	1597.61

Ergebnisse der Parameteroptimierung, Szenario "OPERNRING", $t_{prog} = 0$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	7.47	7.71	2.52	3.84	63	9.16	1823.51
2	10.00	10.00	1.00	4.00	301	10.00	1813.69
3	8.52	4.50	1.00	3.84	63	9.16	1742.03

Ergebnisse der Parameteroptimierung, Szenario "OPERNRING", $t_{prog} = 15$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	9.03	8.72	3.18	1.53	154	5.02	1819.76
3	9.03	8.60	2.25	1.56	154	5.02	1812.34
7	9.03	8.60	2.25	1.56	128	5.99	1799.10

Ergebnisse der Parameteroptimierung, Szenario "OPERNRING", $t_{prog} = 30$ Minuten

KEPPLER-WEST

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	8.74	1.39	2.36	3.17	71	3.45	2583.20
10	8.74	1.29	2.36	3.17	100	3.45	2583.20

Ergebnisse der Parameteroptimierung, Szenario "KEPPLER-WEST", $t_{prog} = 0$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
9	9.02	8.40	1.30	3.70	32	0.19	2599.87

Ergebnisse der Parameteroptimierung, Szenario "KEPPLER-WEST", $t_{Prog} = 15$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	8.74	1.39	2.36	3.17	71	3.45	2882.71
2	10	1	1	1	301	8.88	2755.59
3	10	1	1	1	301	8.94	2755.11
5	10	1	1	1	301	9.18	2753.34
10	9.43	1.02	1	1.07	298	9.10	2735.59

Ergebnisse der Parameteroptimierung, Szenario "KEPPLER-WEST", $t_{Prog} = 30$ Minuten

INGOLSTÄDTER

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	8.73	1.39	2.36	7.51	71	0.35	4124.52
10	8.73	1.29	2.36	8.32	71	0.35	4099.32

Ergebnisse der Parameteroptimierung, Szenario "INGOLSTÄDTER", $t_{Prog} = 0$ Minuten

Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	9.16	2.18	7.84	6.96	18	0.35	5742.37
2	1	1	1	25	1	10	4440.48

Ergebnisse der Parameteroptimierung, Szenario "INGOLSTÄDTER", $t_{Prog} = 15$ Minuten

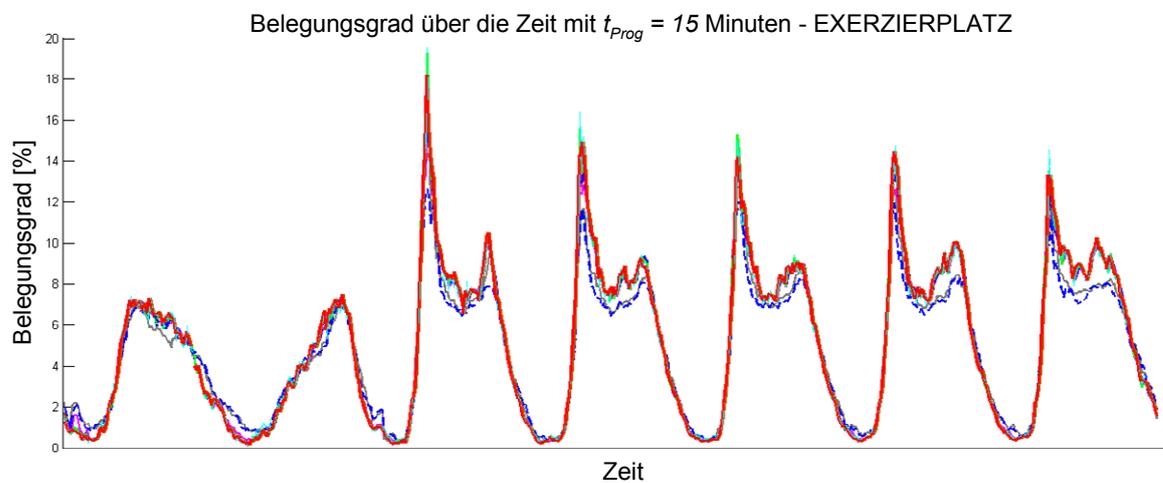
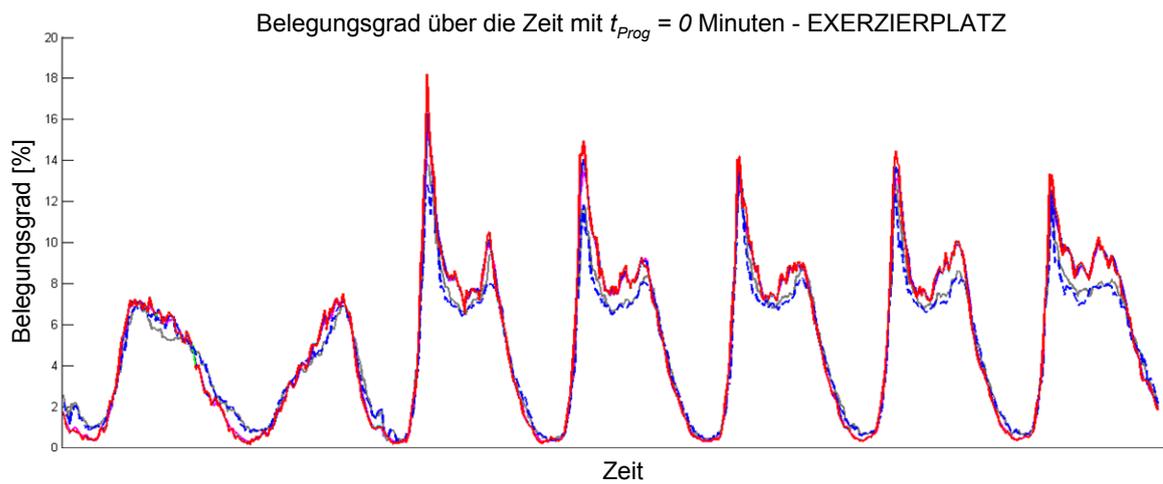
Generation	G_V	k_e	dt_{max}	ds_{max}	k	K_W	MQF
1	8.73	1.39	2.36	7.51	71	0.35	7194.09
2	10	1	1	1	301	0.89	6536.25
5	10	1	1	1	301	0.92	6527.36
6	10	1	10	10	301	1	6504.57
10	10	1	6.18	5.87	290	1	6502.30

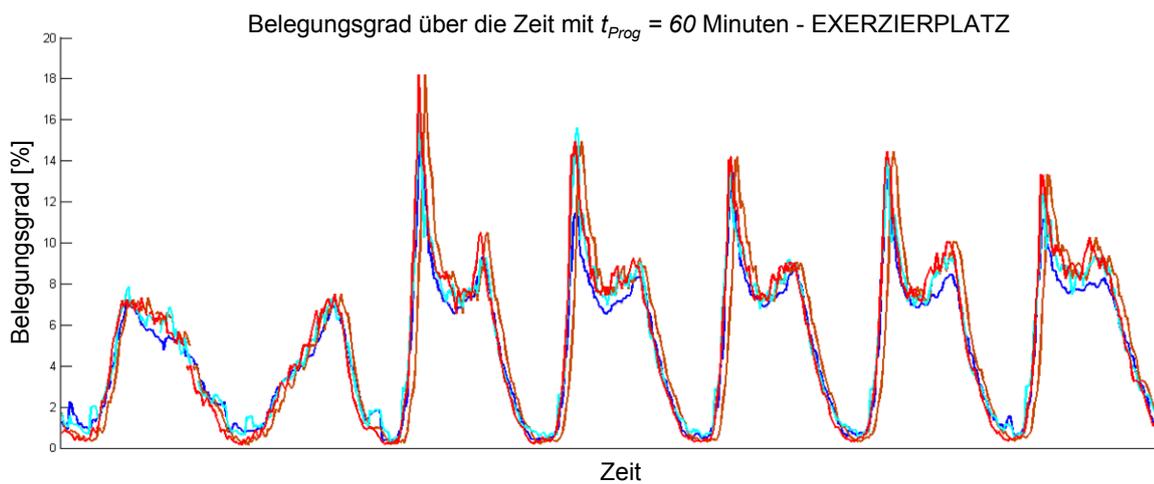
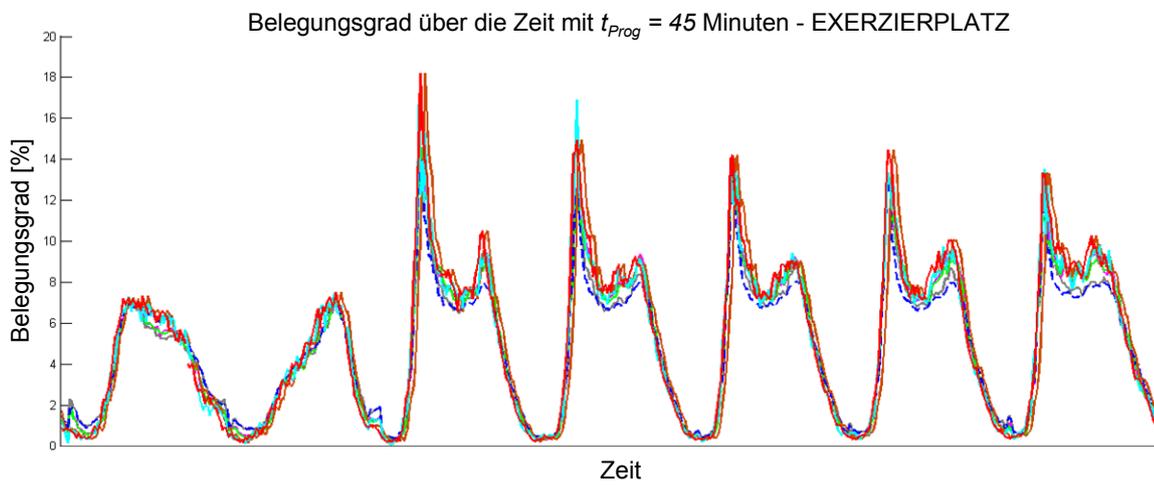
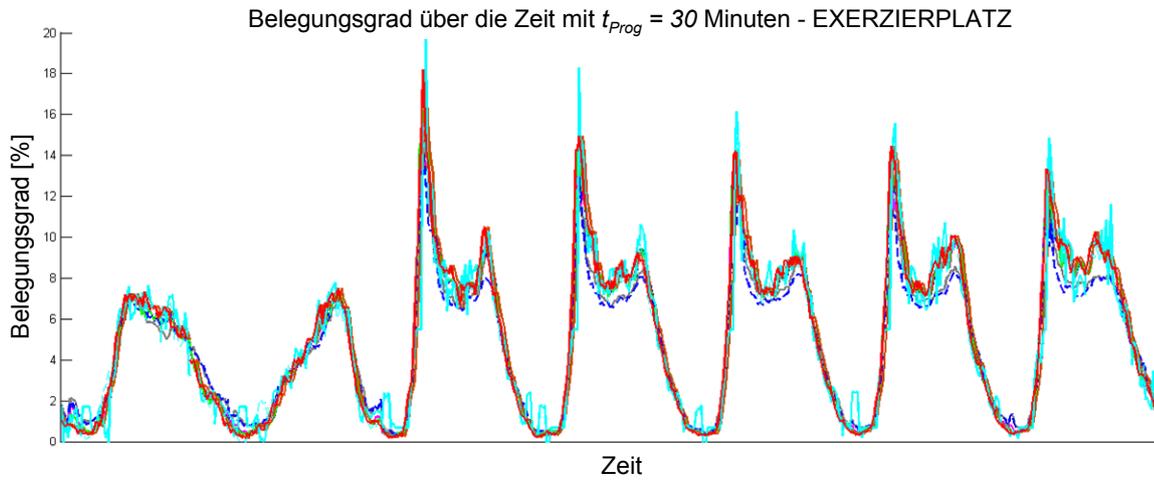
Ergebnisse der Parameteroptimierung, Szenario "INGOLSTÄDTER", $t_{Prog} = 30$ Minuten

Anhang 3 Darstellung der Prognoseergebnisse

EXERZIERPLATZ 1.-7. Dezember 2007, Zeitreihendarstellung

Legende:



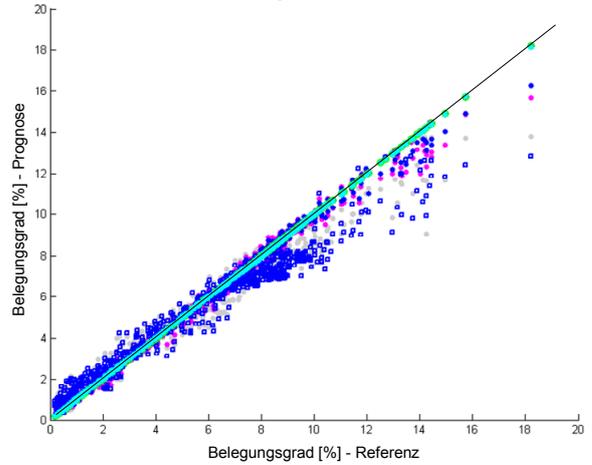


EXERZIERPLATZ 1.-7. Dezember 2007, Scatter-Plots

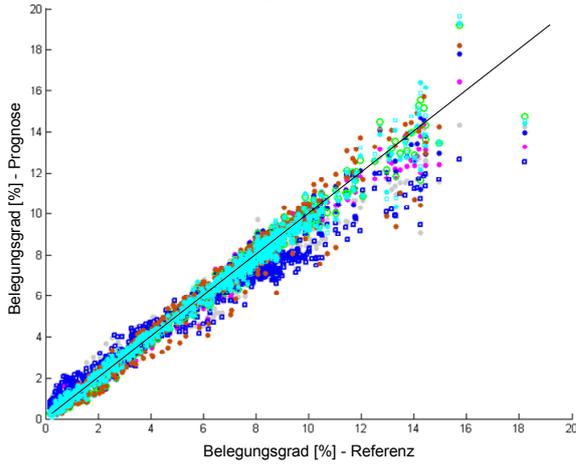
Legende:

- Kalenderbasierte Prognose
- Trivialprognose
- Multilayer Feed Forward
- Multilayer Feed Forward „ohne lokale Messung“
- Instanzbasiertes Lernverfahren
- Instanzbasiertes Lernverfahren „ohne lokale Messung“
- Instanzbasiertes Lernverfahren „nur lokal“
- Instanzbasiertes Lernverfahren mit GL Angleich „nur lokal“
- Instanzbasiertes Lernverfahren mit GL Angleich

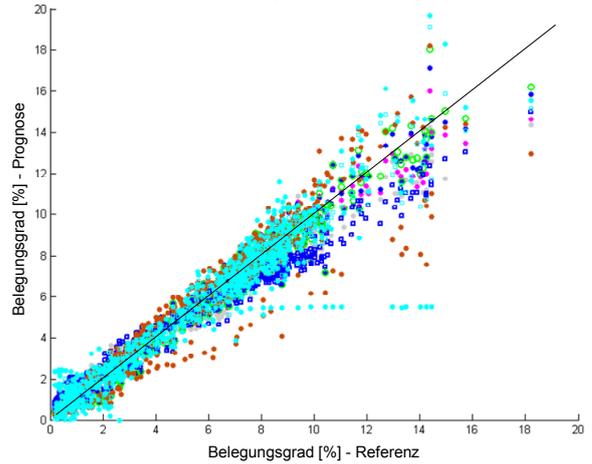
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 0$ Minuten - EXERZIERPLATZ



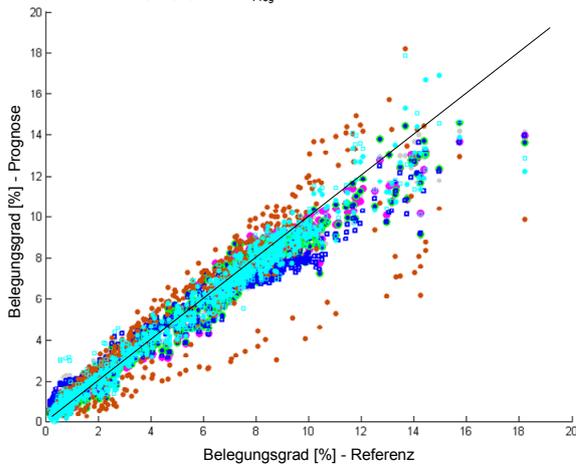
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 15$ Minuten - EXERZIERPLATZ



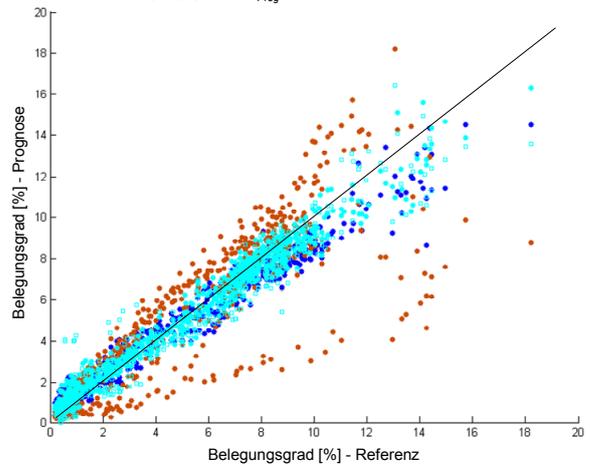
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 30$ Minuten - EXERZIERPLATZ



Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 45$ Minuten - EXERZIERPLATZ

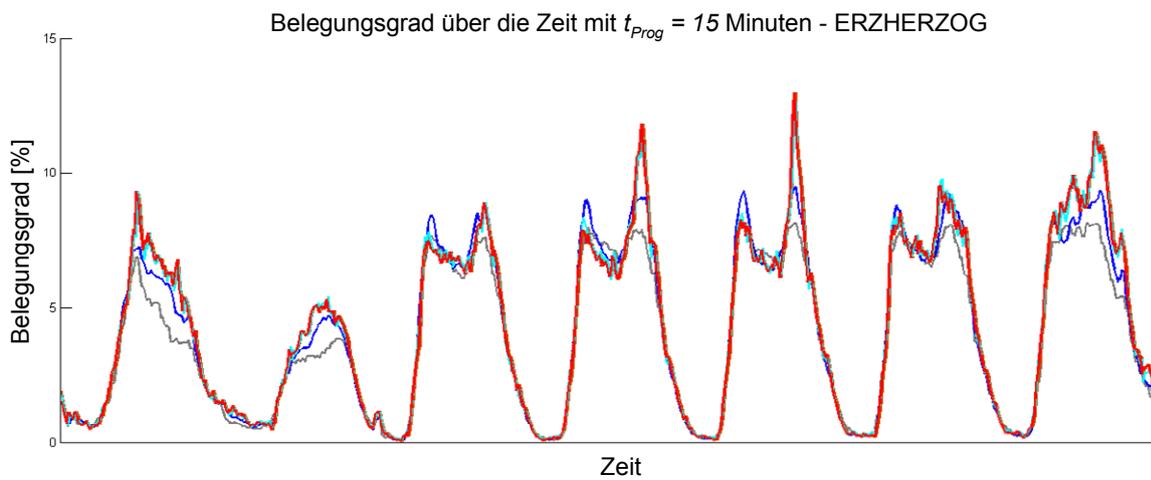
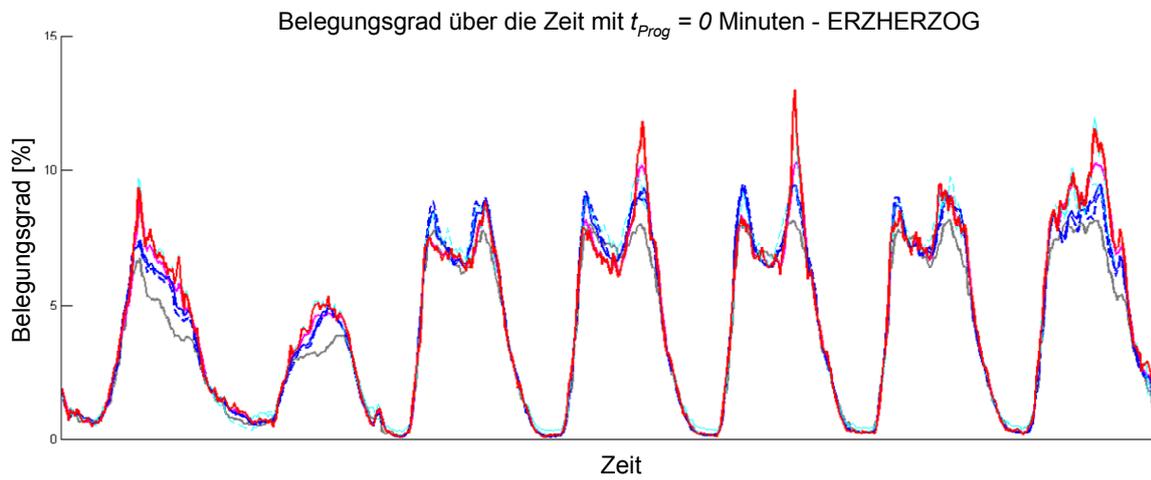


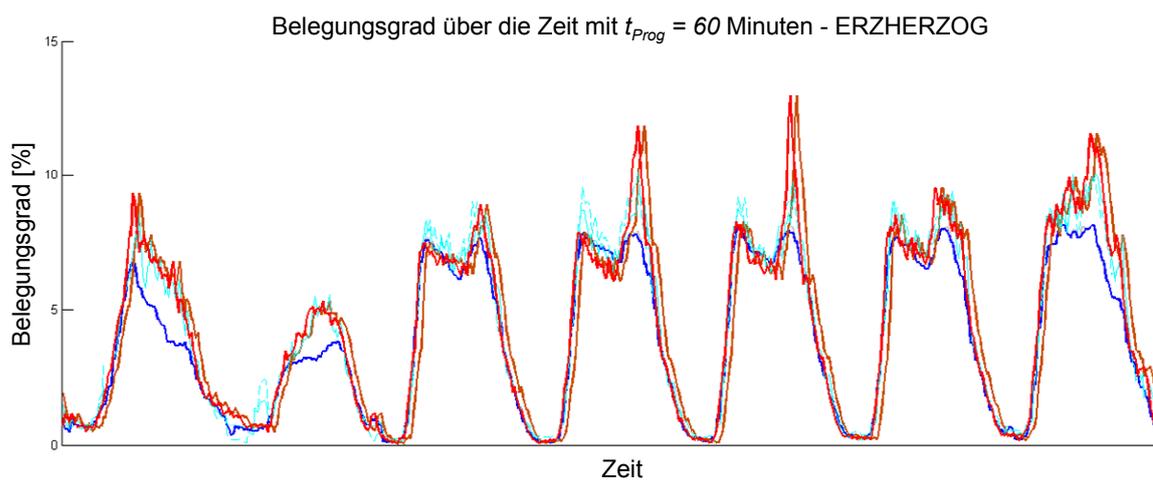
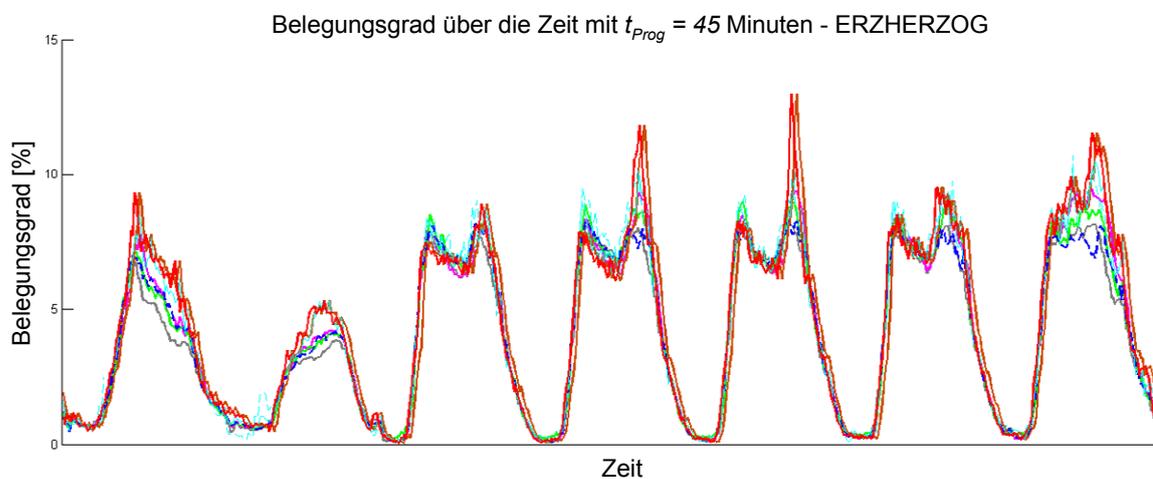
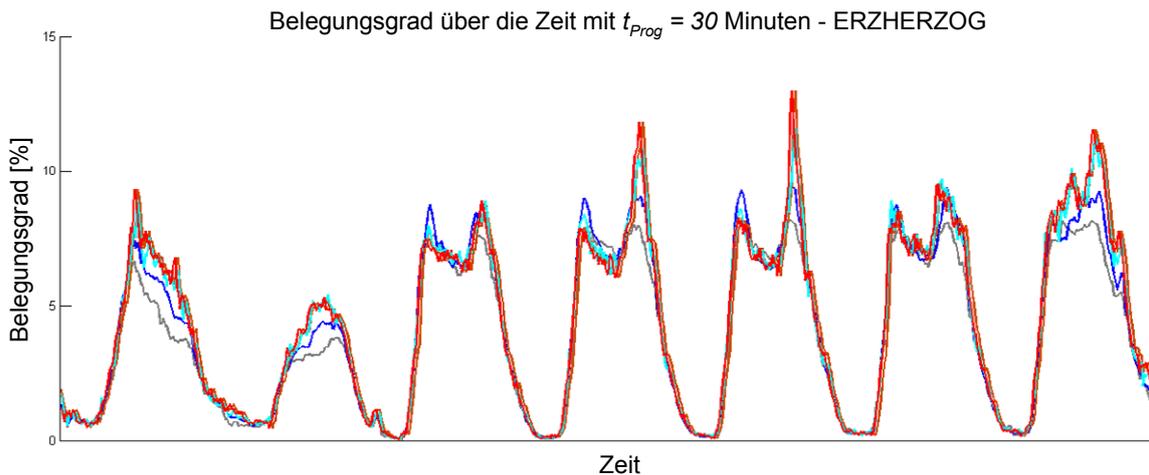
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 60$ Minuten - EXERZIERPLATZ



ERZHERZOG 1.-7. Dezember 2007, Zeitreihendarstellung

Legende:



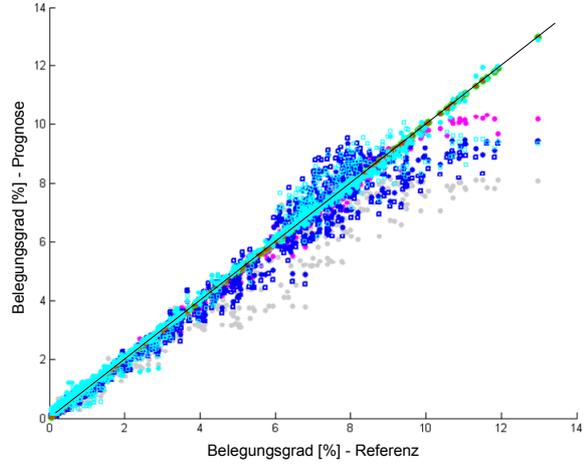


ERZHERZOG 1.-7. Dezember 2007, Scatter-Plots

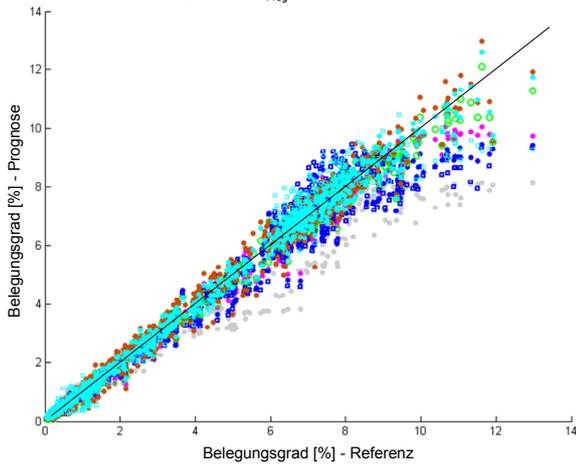
Legende:

- Kalenderbasierte Prognose
- Trivialprognose
- Multilayer Feed Forward
- Multilayer Feed Forward „ohne lokale Messung“
- Instanzbasiertes Lernverfahren
- Instanzbasiertes Lernverfahren „ohne lokale Messung“
- Instanzbasiertes Lernverfahren „nur lokal“
- Instanzbasiertes Lernverfahren mit GL Angleich „nur lokal“
- Instanzbasiertes Lernverfahren mit GL Angleich

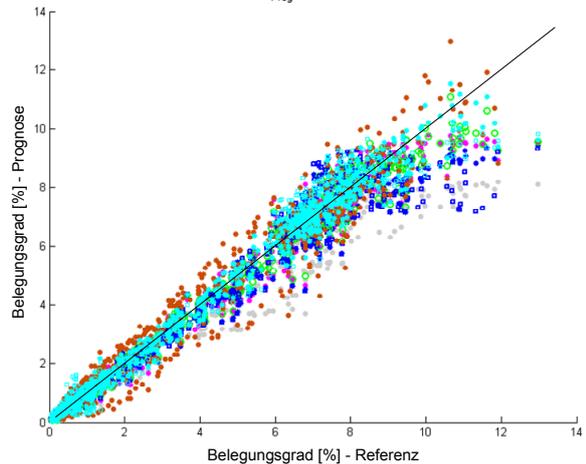
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 0$ Minuten - ERZHERZOG



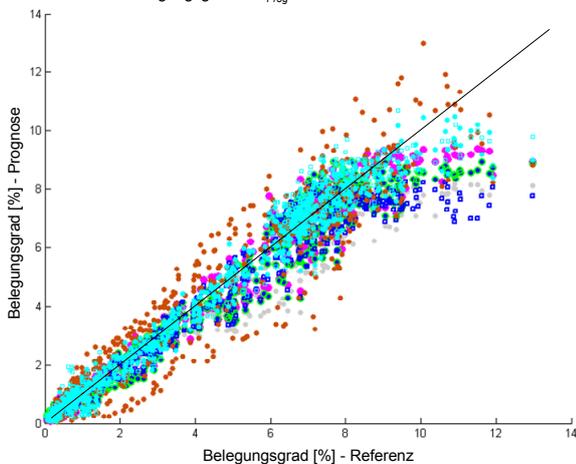
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 15$ Minuten - ERZHERZOG



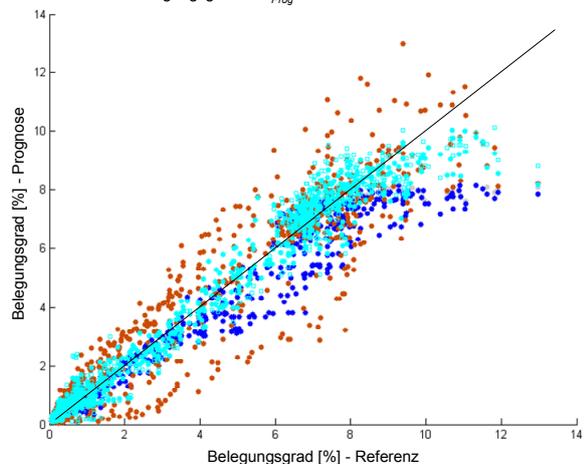
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 30$ Minuten - ERZHERZOG



Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 45$ Minuten - ERZHERZOG

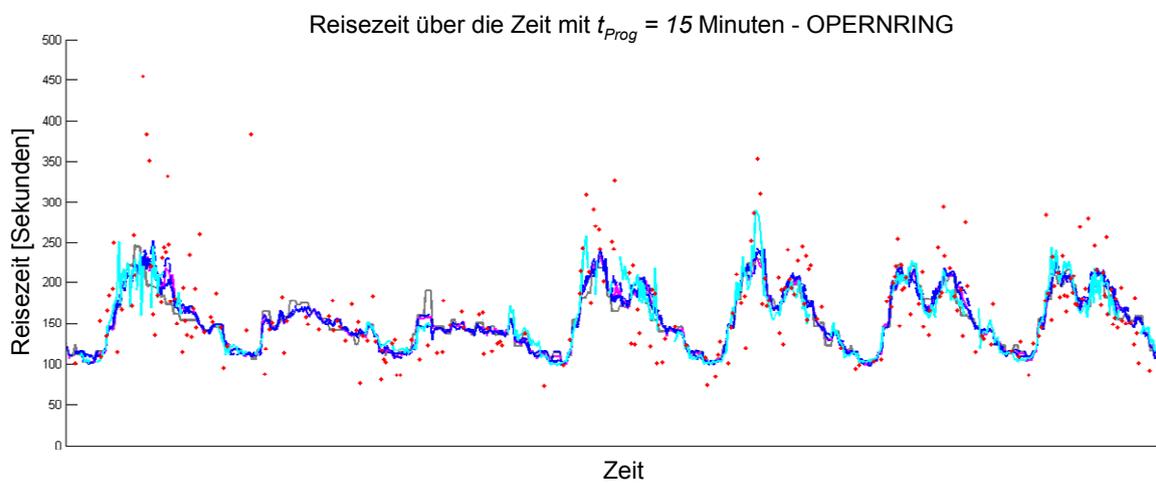
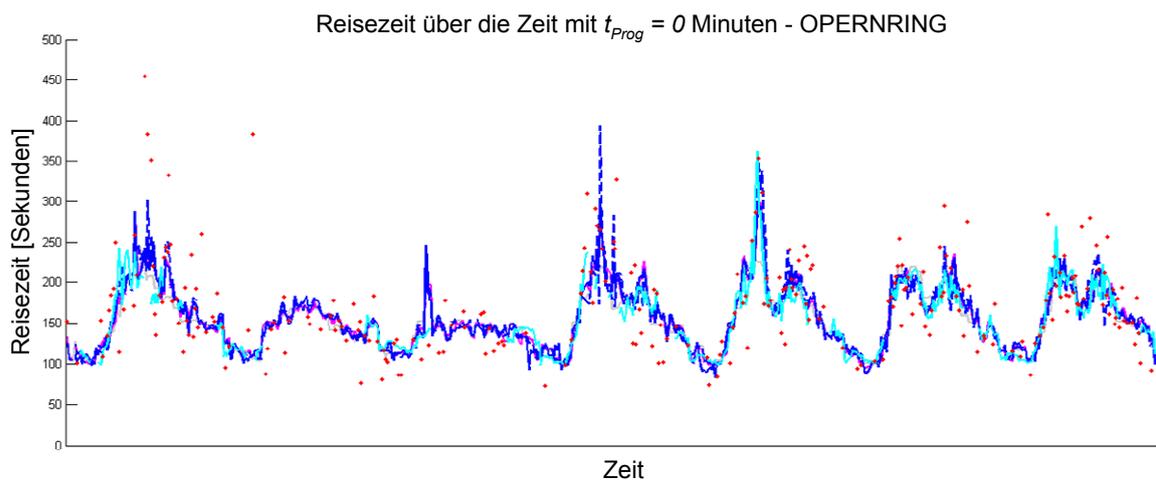
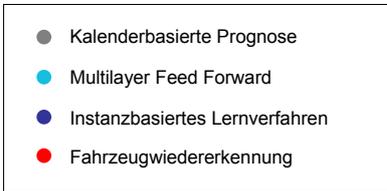


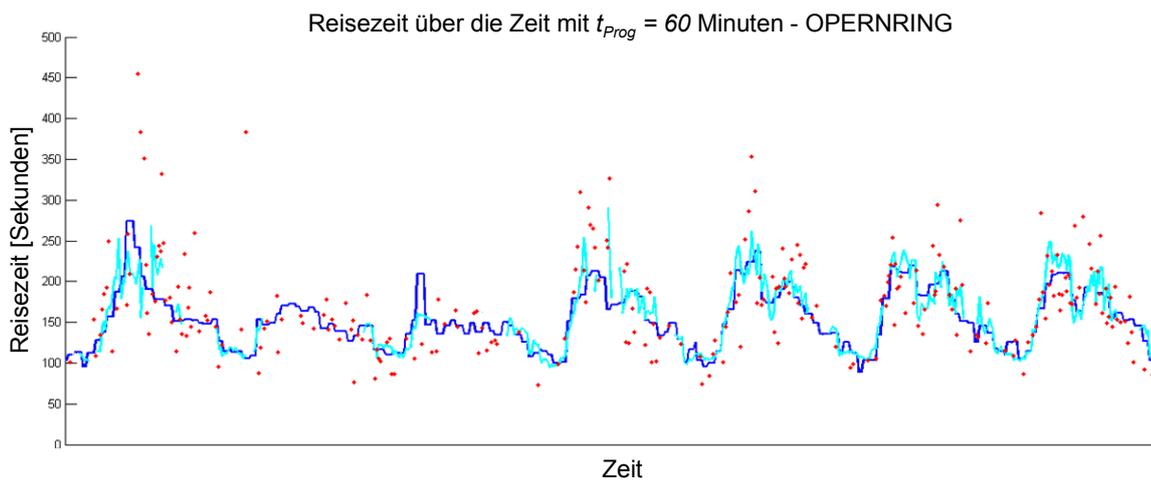
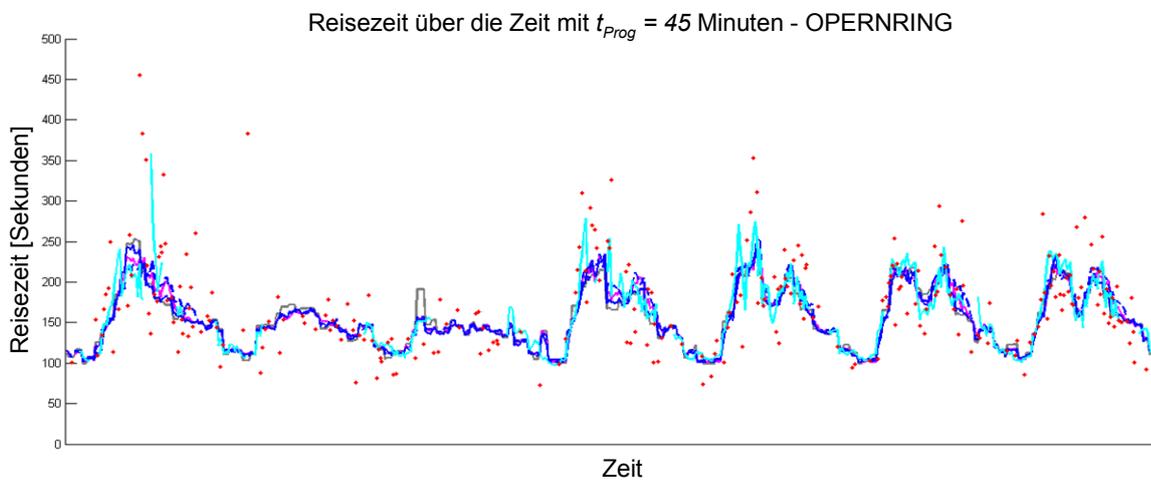
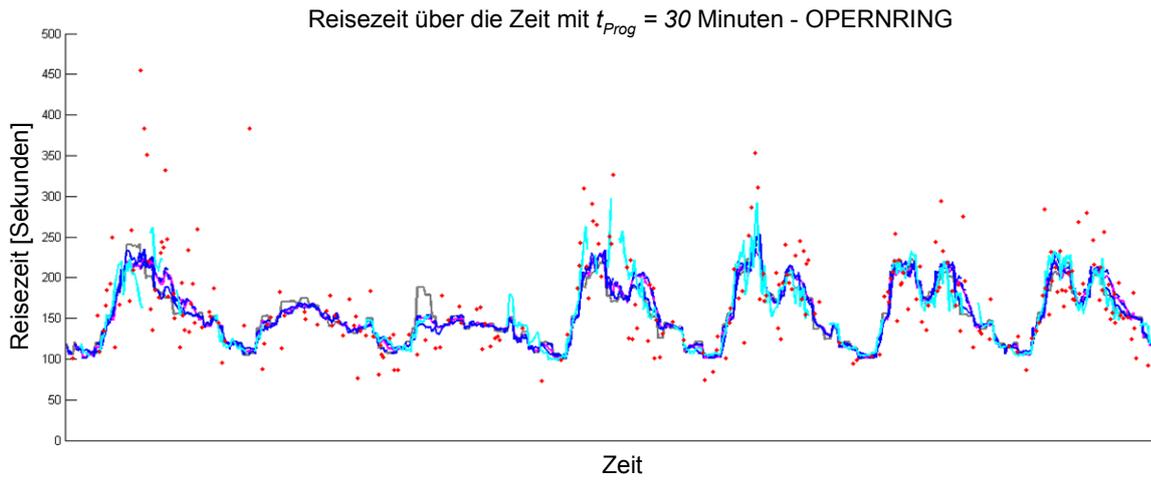
Zusammenhang zwischen Referenzbelegungsgrad und prognostiziertem Belegungsgrad mit $t_{prog} = 60$ Minuten - ERZHERZOG



OPERNRING, 16.-22. November 2007, Zeitreihendarstellung

Legende:



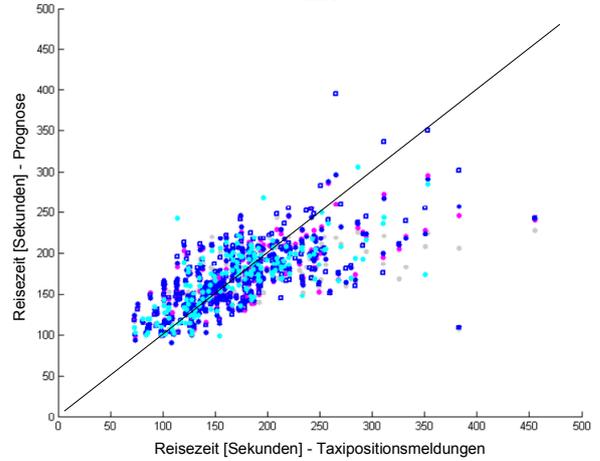


OPERNRING, 16.-22. November 2007, Scatter-Plots

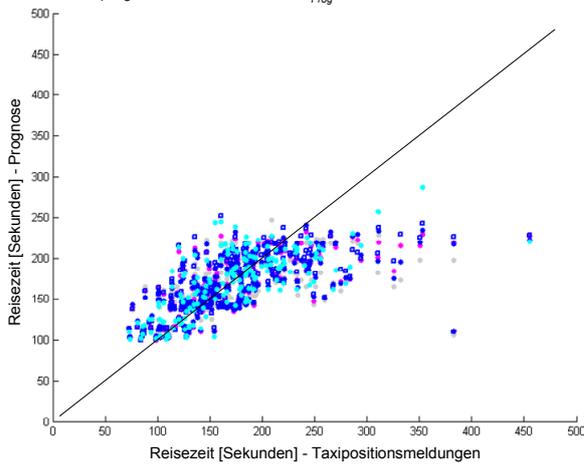
Legende:

- Kalenderbasierte Prognose
- Multilayer Feed Forward
- Instanzbasiertes Lernverfahren
- Instanzbasiertes Lernverfahren „ohne lokale Messung“
- Instanzbasiertes Lernverfahren „nur lokal“

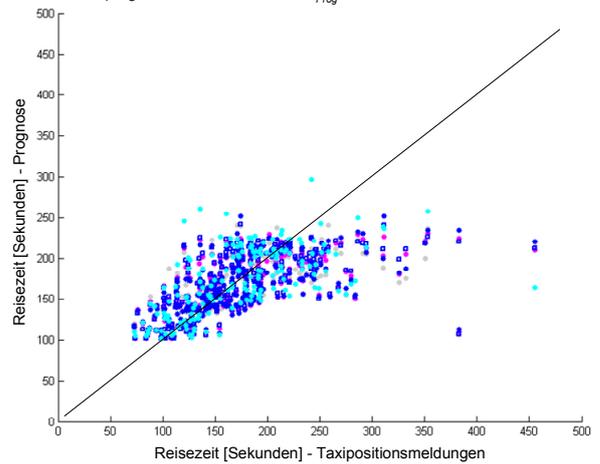
Zusammenhang zwischen Reisezeit aus Taxipositionsmeldungen und prognostizierter Reisezeit mit $t_{prog} = 0$ Minuten - OPERNRING



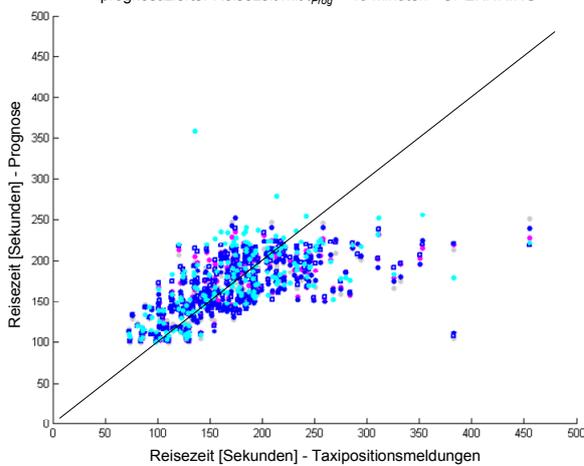
Zusammenhang zwischen Reisezeit aus Taxipositionsmeldungen und prognostizierter Reisezeit mit $t_{prog} = 15$ Minuten - OPERNRING



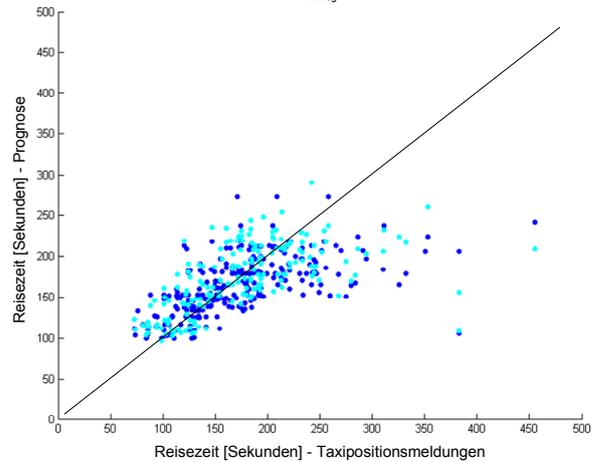
Zusammenhang zwischen Reisezeit aus Taxipositionsmeldungen und prognostizierter Reisezeit mit $t_{prog} = 30$ Minuten - OPERNRING



Zusammenhang zwischen Reisezeit aus Taxipositionsmeldungen und prognostizierter Reisezeit mit $t_{prog} = 45$ Minuten - OPERNRING

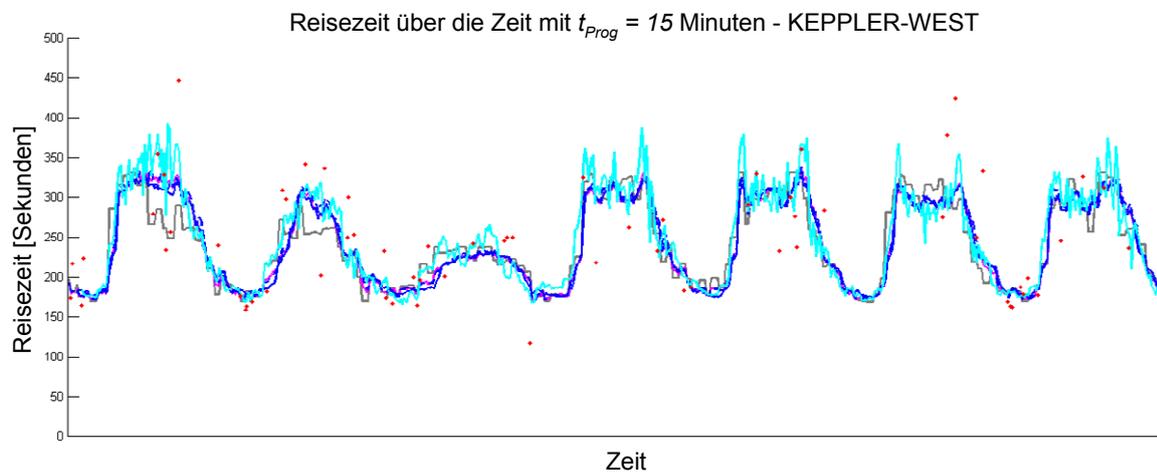
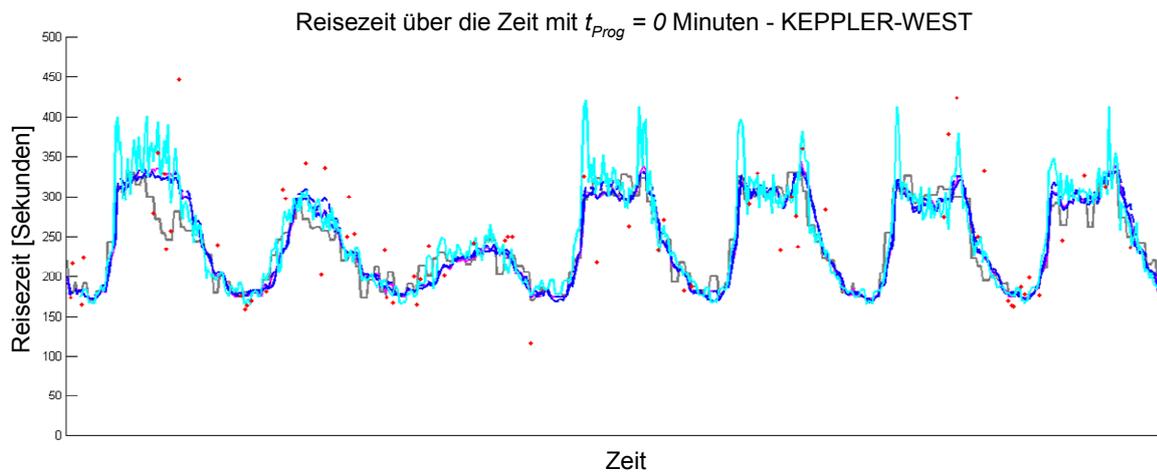


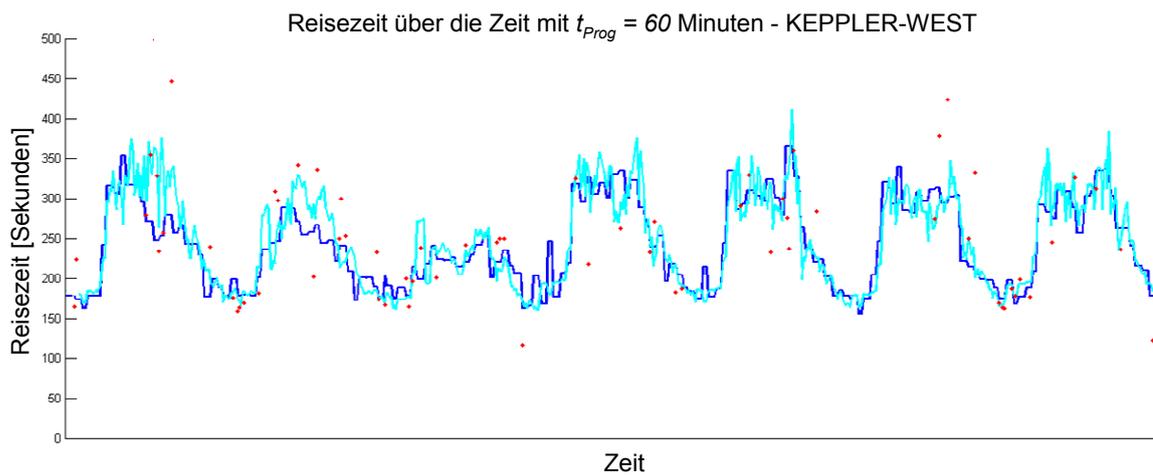
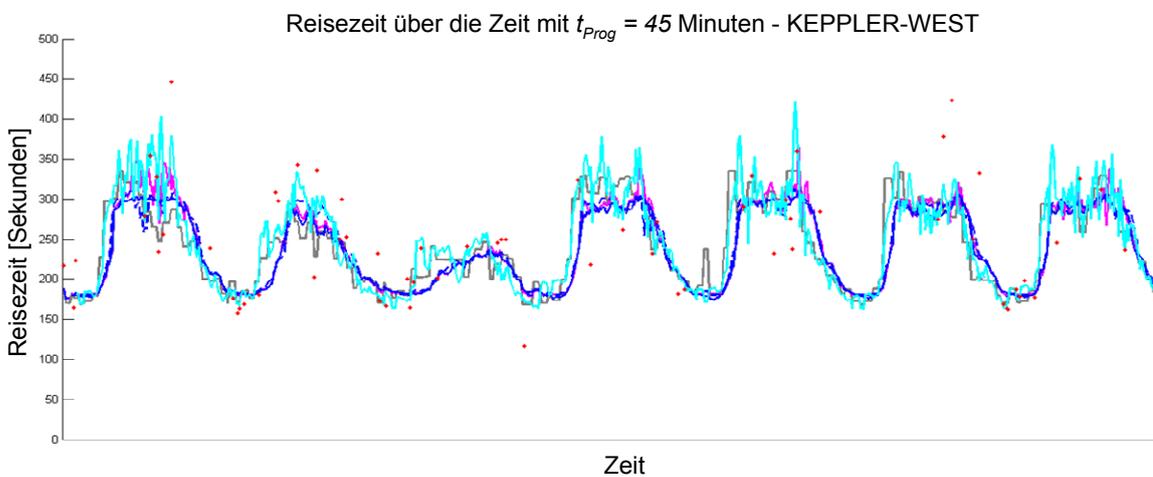
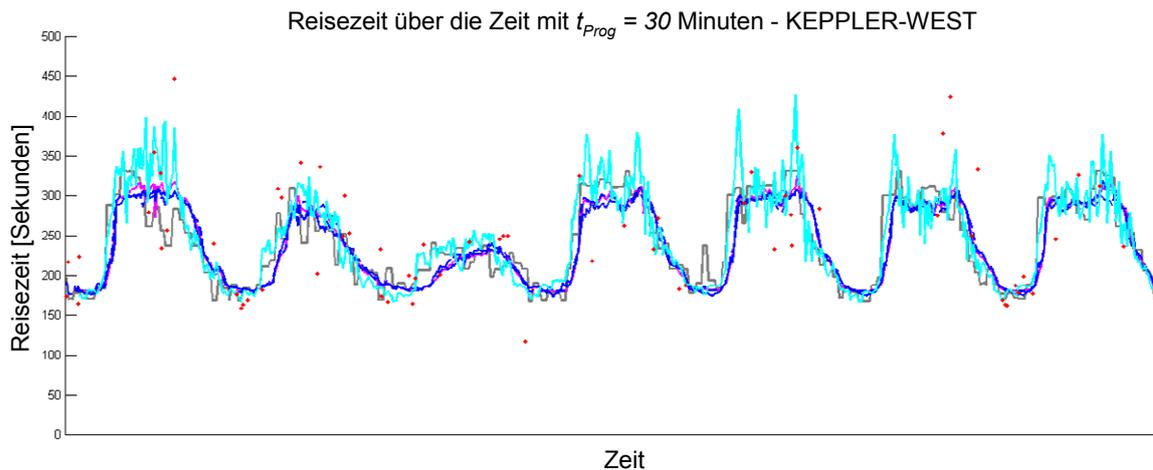
Zusammenhang zwischen Reisezeit aus Taxipositionsmeldungen und prognostizierter Reisezeit mit $t_{prog} = 60$ Minuten - OPERNRING



KEPPLER-WEST, 16.-22. November 2007, Zeitreihendarstellung

Legende:

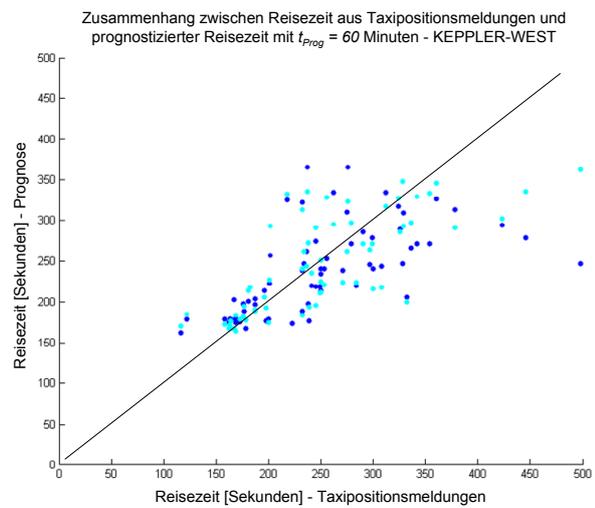
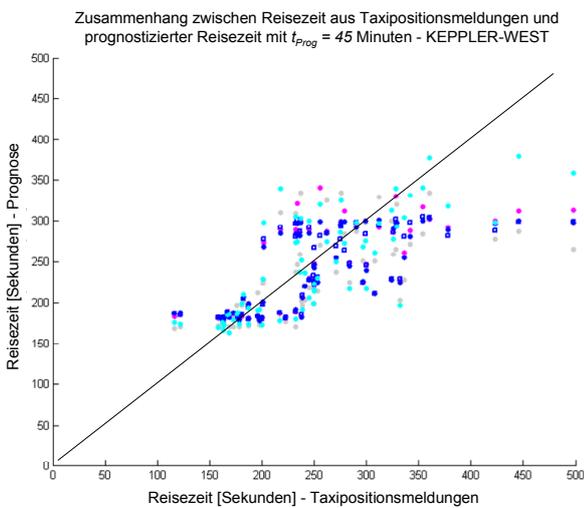
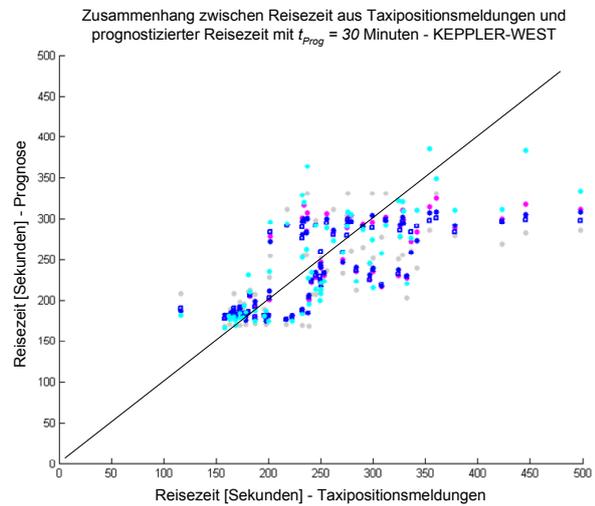
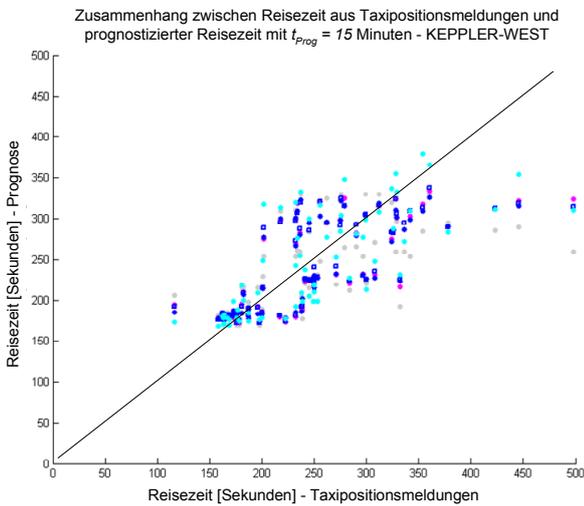
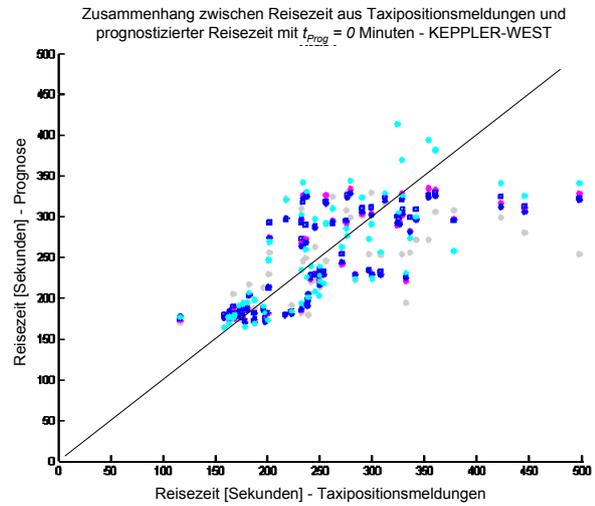




KEPPLER-WEST, 16.-22. November 2007, Scatter-Plots

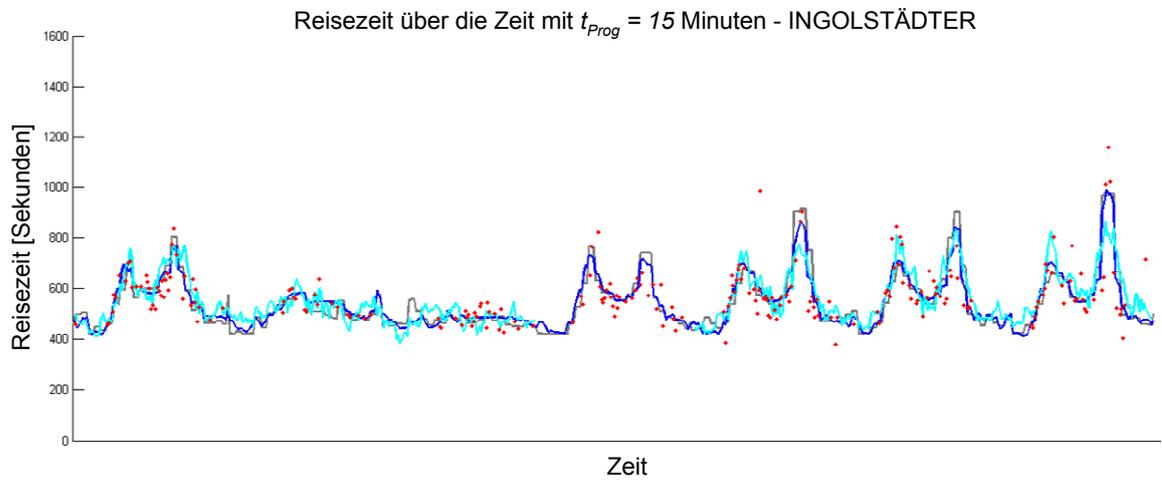
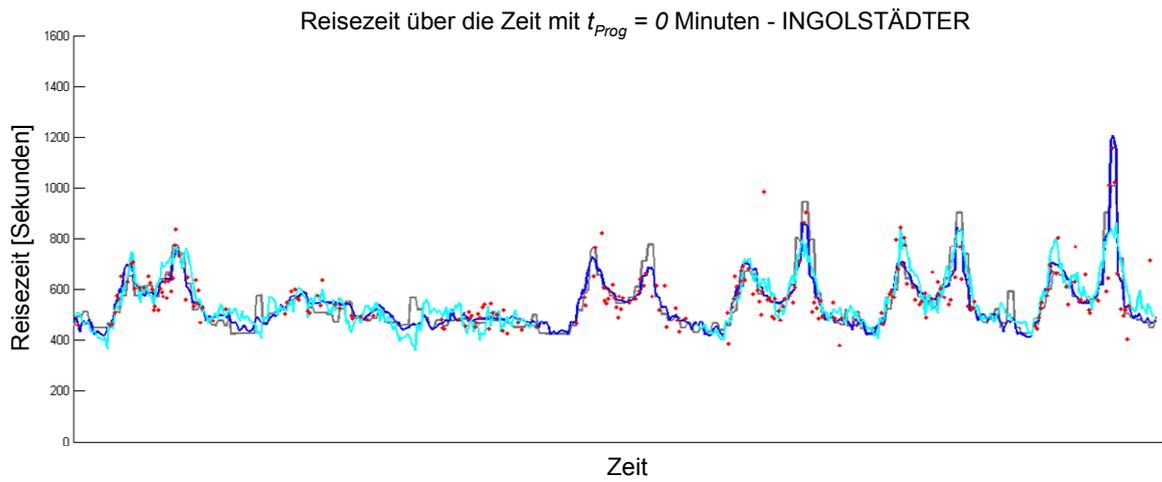
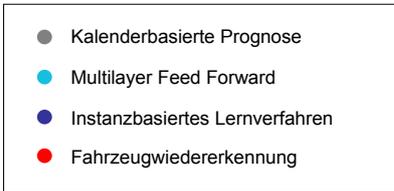
Legende:

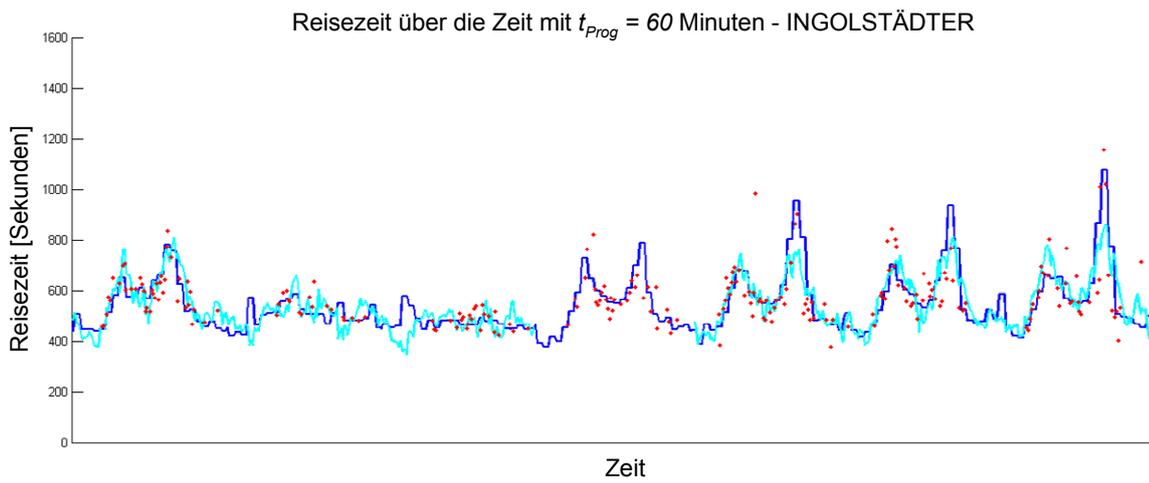
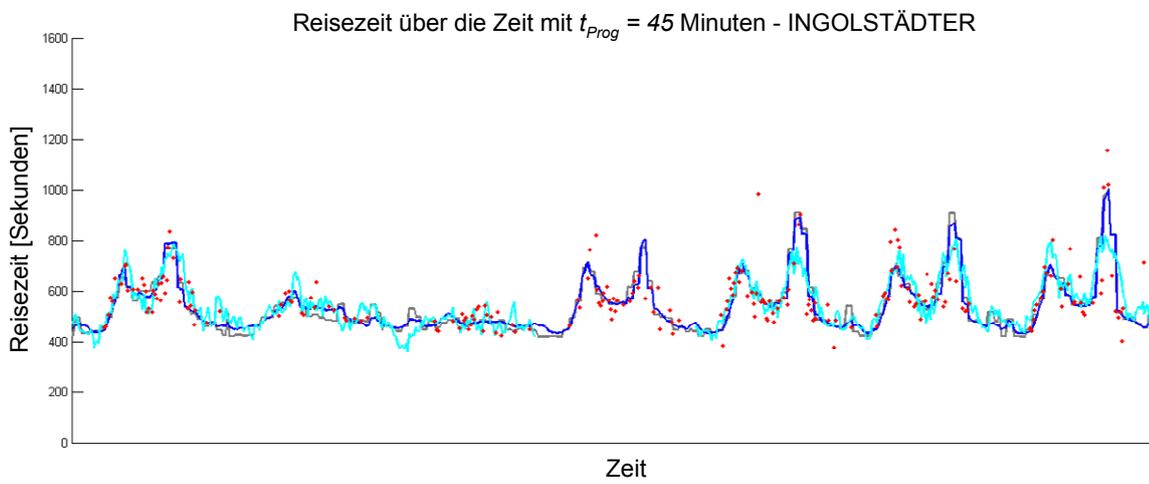
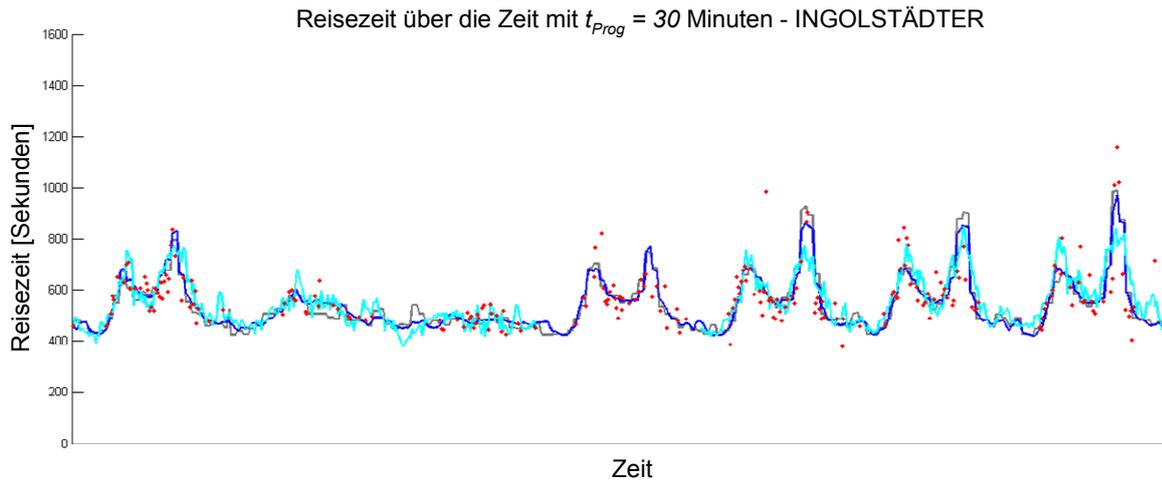
- Kalenderbasierte Prognose
- Multilayer Feed Forward
- Instanzbasiertes Lernverfahren
- Instanzbasiertes Lernverfahren „ohne lokale Messung“
- Instanzbasiertes Lernverfahren „nur lokal“



INGOLSTÄDTER, 16.-22. November 2007, Zeitreihendarstellung

Legende:



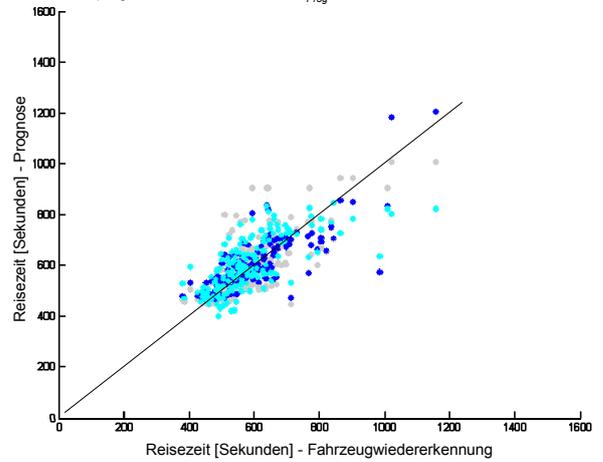


INGOLSTÄDTER, 16.-22. November 2007, Scatter-Plots

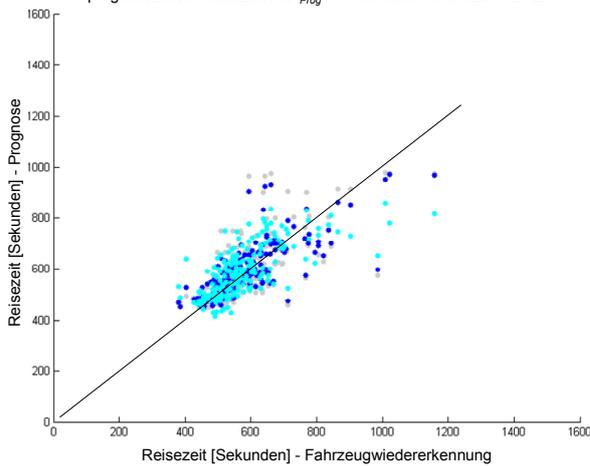
Legende:

- Kalenderbasierte Prognose
- Multilayer Feed Forward
- Instanzbasiertes Lernverfahren

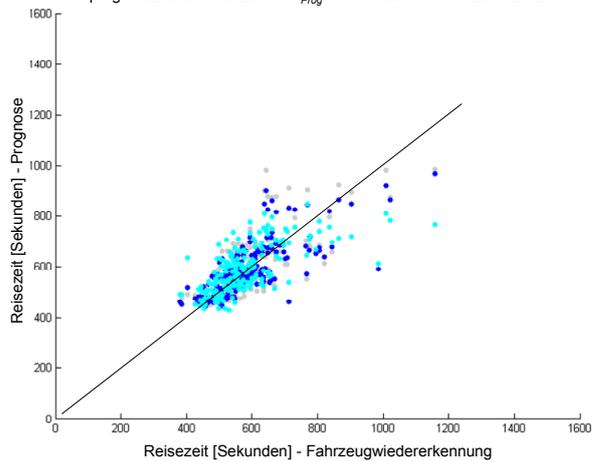
Zusammenhang zwischen Reisezeit aus Fahrzeugwiedererkennung und prognostizierter Reisezeit mit $t_{prog} = 0$ Minuten - INGOLSTÄDTER



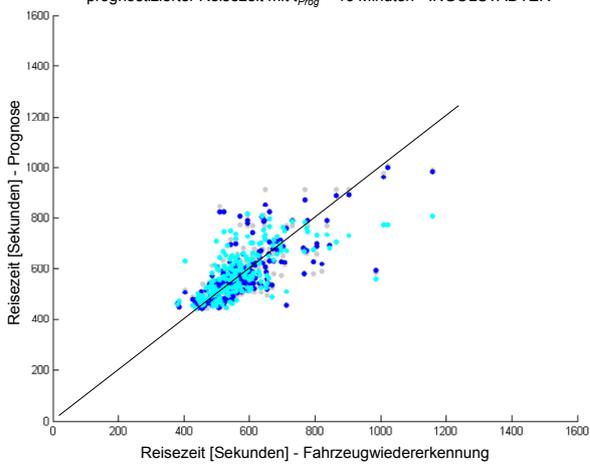
Zusammenhang zwischen Reisezeit aus Fahrzeugwiedererkennung und prognostizierter Reisezeit mit $t_{prog} = 15$ Minuten - INGOLSTÄDTER



Zusammenhang zwischen Reisezeit aus Fahrzeugwiedererkennung und prognostizierter Reisezeit mit $t_{prog} = 30$ Minuten - INGOLSTÄDTER



Zusammenhang zwischen Reisezeit aus Fahrzeugwiedererkennung und prognostizierter Reisezeit mit $t_{prog} = 45$ Minuten - INGOLSTÄDTER



Zusammenhang zwischen Reisezeit aus Fahrzeugwiedererkennung und prognostizierter Reisezeit mit $t_{prog} = 60$ Minuten - INGOLSTÄDTER

