

**Ein Assistenzsystem zur fahrer- und situationsadaptiven Prädiktion
potenzieller Zielorte für eine robuste Interaktion mit sprachgesteuerten
Navigationssystemen**

**An assistance system for driver- and situation-adaptive destination prediction
for a robust interaction with speech controlled navigation systems.**

Autor:

Dipl.-Ing. Marc Hofmann

Lehrstuhl für Mensch-Maschine-Kommunikation

Technische Universität München

D-80290 München

Mail: hofmann@ei.tum.de

Tel.: 089 / 28928543

Co-Autoren:

Dr. Klaus Bengler

BMW AG

Abt. EV-22

D-80788 München

Mail: Klaus.Bengler@bmw.de

Tel.: 089 / 38244999

Prof. Dr. rer. nat. Manfred Lang

Lehrstuhl für Mensch-Maschine-Kommunikation

Technische Universität München

D-80290 München

Mail: lang@ei.tum.de

Tel.: 089 / 28928541

Zusammenfassung:

Im Rahmen dieses Beitrags wird ein Assistenzsystem vorgestellt, dessen Zielsetzung es ist, einen robusteren und effizienteren sprachlichen Dialog mit Navigationssystemen zu ermöglichen. Durch fahrer- und situationsadaptive Gewichtung aller Orte Deutschlands hinsichtlich ihrer Relevanz als potenziellen Zielort, können mehrdeutige Ortsnamen dem Wunsch-Zielort des Fahrers zugeordnet werden, um somit Klärungsdialoge und Systemrückfragen zu minimieren. Diese Informationen werden zudem zur adaptiven Begrenzung des Ortsnamen-Vokabulars des Spracherkenners verwendet, sowie zum intelligenten Anbieten eines Default-Zielorts. Die Zielort-Prädiktion erfolgt aufgrund einer quantitativen Analyse der Korrelation von Zielort und Situation, wobei Aussagen über bisher nicht besuchte Orten und unbekannt Situationen getroffen werden können.

Abstract:

We present a navigation assistance system for a more robust and efficient speech-based dialog with navigation systems. Taking driver and situation adaptive criteria of all German locations into account regarding their relevance as potential destination, ambiguous location names can be assigned to the desired destination for reducing system requests. This information is also used for an adaptive limiting of the speech recognizer's location-vocabulary, as well as for an intelligent offering of the default destination. The destination prediction is based on a quantitative analysis of the correlation of destination and situation, with the system being able to reason about not attended locations and unknown situations.

1. Einleitung

Sprachsteuerung etabliert sich gegenwärtig für kompakte, abgeschlossene Applikationen mit dem Anspruch einer intuitiv bedienbaren Schnittstelle zu informationstechnischen Systemen. Für den Einsatz von Spracherkennung im Fahrzeug ist neben dem Vorteil der intuitiven Bedienbarkeit und des Komforts in erster Linie das Argument ausschlaggebend, dass Sprachsteuerung, verglichen mit haptischer Eingabe, die Ablenkung des Fahrers minimiert. Vor allem für Fahrzeug-Navigationssysteme birgt Sprachsteuerung großes Potenzial. Dem stehen im Moment noch zwei Probleme entgegen:

Ein Problem bei der sprachlichen Eingabe eines Ortsnamens besteht in der Tatsache, dass der Fahrer oft den korrekten, offiziellen Ortsnamen nicht kennt bzw. nicht spricht. In der Regel wird der Fahrer beispielsweise "Neustadt" als Ziel angeben und nicht die vollständige Bezeichnung "Neustadt an der Aisch". Aus der Mehrdeutigkeit einer Vielzahl von Ortsnamen ergibt sich somit das Problem, dass das Navigationssystem nicht in der Lage ist, die interpretierte Information dem richtigen Ort zuzuweisen. Aktuelle Navigationssysteme können diese Ambiguitäten nur anhand von Klärungsdialogen auflösen. Kognitiv nicht unmittelbar erfassbare, komplexe Systemrückfragen erhöhen jedoch die Ablenkung des Fahrers und reduzieren den Komfort und die Freude am Umgang mit dem entsprechenden System.

Ein weiteres Problem liegt in der Tatsache begründet, dass das Spracherkenner-Vokabular für die sprachliche Eingabe des Zielorts allein für Deutschland ca. 80000 Einträge umfasst. Proportional zu dieser Zahl ist die Fehlerrate des Spracherkenners, d.h. bei dem derzeitigen Stand der Technik müsste der Fahrer rein statistisch den Zielort dem System mehrfach per Sprache mitteilen, bevor der Zielort korrekt erkannt wird. Im Fall einer Häufung einer solchen Situation würde die Effizienz der sprachlichen Schnittstelle und somit die Akzeptanz des

gesamten Systems reduziert. Da unvorhersehbares Verhalten eines informationstechnischen Systems für zusätzliche Ablenkung sorgt, können Fehlerkennungen zudem die Sicherheit beeinträchtigen.

2. Motivation

Für eine robuste und komfortable Interaktion mit Navigationssystemen mittels Sprache ist die Reduzierung der erwähnten Phänomene unumgänglich. Beide Probleme könnten durch die Fähigkeit des Navigationssystems, Aussagen über potenzielle Zielorte zu treffen, minimiert werden. Ziel des hier vorgestellten Verfahrens ist die Prädiktion des Wunsch-Zielorts des Fahrers. Der Begriff der Prädiktion wird im Rahmen dieses Verfahrens als eine Gewichtung aller Orte hinsichtlich ihrer Relevanz als potentielle Zielorte für den Fahrer, interpretiert. Diese Zielort-Prädiktion wird schließlich verwendet, um die Kommunikation mit dem Navigationssystem hinsichtlich folgender Aspekte zu verbessern:

- Disambiguierung von Zielorteingaben: Die Fähigkeit des Systems, Zielort-Eingaben durch Evaluierung und Gewichtung aller Orte Deutschlands zu interpretieren, soll zur korrekten Zuordnung mehrdeutiger Ortsnamen verwendet werden, um somit Klärungsdialoge und Systemfragen zu minimieren bzw. zu vermeiden.
- Begrenzung, bzw. Gewichtung des Ortsnamen-Vokabulars: Für Fahrer, die im allgemeinen nur in bestimmten Gegenden agieren, wird dem Spracherkenner ein entsprechend reduziertes Vokabular angeboten um die Erkennungsleistung des Spracherkenners zu erhöhen. Zudem ermöglicht das Verfahren die Nach-Evaluierung und Interpretation des Erkenners-Ergebnisses. Dadurch wird das Navigationssystem in die Lage versetzt, die Zielort-Eingabe nicht nur auf Basis akustischer Analysen zu erkennen, sondern auch inhaltlich zu interpretieren, um somit das Erkennungs-Ergebnis auf Plausibilität zu untersuchen und gegebenenfalls andere Erkenners-Hypothesen vorzuziehen.
- Aktuelle Navigationssysteme bringen folgende paradoxe Situation mit sich: Startet der Fahrer das Navigationssystem, so wird ihm als Default-Zielort der zuletzt eingegebene Zielort angeboten. Im allgemeinen handelt es sich dabei allerdings um den aktuellen Standort. Um diese paradoxe Situation aufzulösen ist das dritte Ziel unseres Verfahrens eine intelligente, adaptive Wahl des Default-Zielorts, so dass der Fahrer lediglich bestätigen muss oder gegebenenfalls einen anderen Zielort eingeben kann. Die Zielgruppe für diese Systemeigenschaft sind zum Beispiel Pendler, die in entsprechenden Situationen regelmäßig die gleichen Orte aufsuchen.

Das übergeordnete Ziel unseres Verfahrens ist eine Reduzierung von Klärungsdialogen und von Fehlerkennungen mittels Zielort-Prädiktion. Die resultierende schnelle Konvergenz des Mensch-Maschine-Dialogs führt zu einer benutzerfreundlicheren, robusteren sprachlichen und damit intuitiveren Interaktion mit Navigationssystemen. Im folgenden Abschnitt wird nun auf das hierfür entwickelte Navigationsassistenzsystem *Adaptive Compass* eingegangen.

3. Das Navigationsassistenzsystem *Adaptive Compass*

3.1. Grundidee

Die Grundidee von *Adaptive Compass* wird am besten durch die Metapher des „digitalen, mobilen Sekretariats“ umschrieben: Eine Sekretärin kann auf Grund der Termine auf den Zielort des Fahrers schließen; im Laufe der Zeit lernt die Sekretärin, welche Art von Ziel für den Fahrer in bestimmten Situationen interessant ist und weshalb gerade diese Ziele bevorzugt werden. Die menschliche Fähigkeit, aus Beobachtungen zu lernen, ist die Motivation, das Assistenzsystem mit entsprechenden Fähigkeiten auszustatten, um Zielort-Vorhersagen auf Basis unterschiedlichster Informationen und Informationsquellen zu treffen. Diese Informationen deklarieren wir zunächst pauschal als Situation; in Abschnitt 3.3 folgen nähere Erläuterungen zur Interpretation einer Situation im Rahmen von *Adaptive Compass*. Mittelpunkt des hier vorgestellten Navigationsassistenzsystems ist das Korrelationsmodell, ein Expertensystem zur Modellierung einer potentiellen Korrelation zwischen Zielorten und Situationen. Die Korrelationsfunktion f beschreibt fahrerspezifische und situative Muster bei der Wahl von Zielorten:

$$\text{Zielort} = f(\text{Fahrer}, \text{Situation}) \quad (1)$$

Wie Abb. 1 zeigt, wird das Korrelationsmodell in zwei Modi betrieben, dem Trainingsmodus und dem Inferenzmodus.

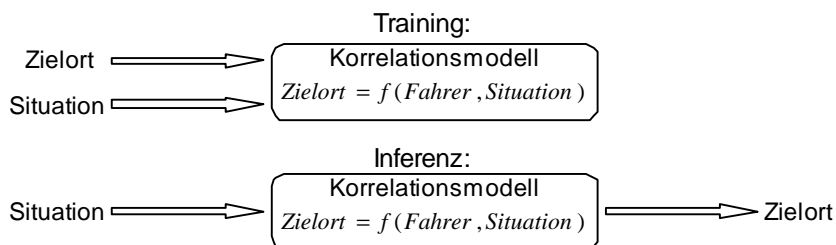


Abb. 1: Trainings- und Inferenzmodus des Korrelationsmodells

Fig.1: Training and inference-mode of the correlation model

Bei einem neu eingegebenen Zielort wird im Trainingsmodus die Korrelation zwischen Situation und Zielort quantitativ erfasst, d.h. der Trainingsmodus dient zur Adaption der Korrelationsfunktion f an eine neue Konstellation von Zielort und Situation. Der

Inferenzmodus wird dient dazu, unter Verwendung der Korrelation f fahrer- und situationsadaptive Aussagen über den potentiellen Zielort zu treffen.

3.2 Anforderungen

Die im Abschnitt 3.1 erwähnte Korrelationsfunktion f stellt nur eine starke Vereinfachung der Grundidee des Verfahrens dar. Eine direkte Korrelation zwischen Zielort und Situation im Trainings-Modus zu berechnen und im Inferenz-Modus als Basis für eine Zielort-Prädiktion zu verwenden ist allerdings in der Praxis nicht sinnvoll. Zum einen wäre der Navigationsassistent dann nur in der Lage, über bereits angefahrte Orte zu urteilen. Zum anderen wären Prädiktionen ausschließlich für bekannte Situationen möglich, d.h. viel zu lange Trainingsphasen wären nötig, um alle potentiellen Situationen zu berücksichtigen.

Daraus ergeben sich folgende Anforderungen an *Adaptive Compass*:

- Urteilen über bisher unbekannte Situationen:
Durch Modellierung der Korrelation in Abhängigkeit jedes einzelnen Situationsparameters erhält *Adaptive Compass* die Fähigkeit, Zielort-Prädiktionen auch für unbekannte bzw. noch nicht exakt beobachtete Situationen vorzunehmen.
- Urteilen über vom Fahrer noch nicht aufgesuchte Orte:
Durch Modellierung der Korrelation zwischen Situation und bestimmten charakteristischen Eigenschaften eines Ortes erhält *Adaptive Compass* die Fähigkeit, über noch nicht aufgesuchte Orte, mit entsprechenden Eigenschaften, Aussagen zu treffen.

Zu beiden Punkten folgen in den nächsten Abschnitten nähere Erläuterungen.

3.3 Definition des Situationsbegriffs

Im Rahmen des Navigationsassistenzsystems umfasst der Situationsbegriff unterschiedlichste Parameter aus den Bereichen Gesamtfahrzeug, Fahrzeugumgebung und den informationstechnischen Einrichtungen des Fahrzeugs.

Die Beschreibung einer Situation erfolgt anhand des Zustandsvektors \underline{x}_{Sit} , der aus einer Reihe von diskreten Zustandsvariablen besteht. Bei der Auswahl der Situationsparameter ist dabei ein Kompromiss zwischen detaillierter Beschreibung der Situation und Einschränkung der Komplexität zu treffen. Der Zustandsvektor \underline{x}_{Sit} ist also als Näherung an die "echte" Situation zu sehen, zumal die Verfügbarkeit von Informationen bezüglich einer Situation unter normalen Umständen bereits zur Abstraktion zwingt. Eine Situationsinformation wird auf eine Bool'sche Zustandsvariable abgebildet um nicht rivalisierende Ereignisse zu

modellieren. Es gibt also nicht eine Zustandsvariable „Tag“, die alle Tage als Zustandsraum besitzt, sondern für jeden Tag eine eigene Zustandsvariable. Diese Einteilung ist auf die Modellierung des Korrelationsmodells zurückzuführen; Abschnitt 3.5.1 wird dies verdeutlichen. Als Situationsparameter sind dabei zwei Arten von Informationen interessant. Zum einen Parameter, die möglicherweise direkten Einfluss auf eine Zielauswahl haben könnten; zum Beispiel könnte schönes Wetter den Fahrer dazu veranlassen, einen See oder die Berge zu besuchen. Zum anderen Parameter, die als Begleiterscheinung bestimmter Ziel-Interessen zu interpretieren sind; zum Beispiel ist zu erwarten, dass während der Arbeitszeit andere Ziele als am Wochenende bevorzugt werden. Situationsparameter, die keiner der beiden Gruppen angehören, eignen sich nicht zur Klassifikation und bleiben deshalb unberücksichtigt.

3.4 Ausgewertete Orts- und Fahrerinformationen

3.4.1 Ortspezifische Informationen

Einen Zusammenhang zwischen Situationen und Orten herzustellen ist nur dann sinnvoll, wenn das Verfahren in der Lage ist, Aussagen über bisher noch nicht angefahrene Orte zu treffen. Um dies zu ermöglichen, muss das Korrelationsmodell die Fähigkeit besitzen, durch Analyse der bisherigen Zielorte, zu ermitteln, welche charakteristischen Eigenschaften einer Gegend für den Fahrer in der jeweiligen Situation interessant sind. Hierfür wurden alle Orte Deutschlands hinsichtlich ihres wirtschaftlichen Potentials und hinsichtlich ihres Freizeitwerts evaluiert. Die wirtschaftlichen Kenngrößen gestatten Aussagen über berufliche Interessen des Fahrers; durch Tourismus-Kenngrößen können Präferenzen des Fahrers in der Freizeit erfasst werden. Die Charakterisierung der Orte erfolgt auf Kreisebene, da Daten des Statistischen Bundesamts verwendet wurden, die Deutschland nur auf Kreisebene beschreiben und sich zudem die Komplexität der Datenbanken in Grenzen halten lässt.

Anhand der Daten wurde berechnet, inwiefern bestimmte Industriezweige für die einzelnen Kreise charakteristisch sind. Hierzu wurde zunächst der Anteil der Erwerbstätigen in einer bestimmten Branche KG (Kenngröße) für den Kreis bestimmt. Für einen beliebigen Kreis geschieht dies durch Division der Anzahl der Erwerbstätigen in der jeweiligen Branche $n_{KG_{Kreis}}$ durch die Gesamtzahl der Erwerbstätigen $n_{Gesamt_{Kreis}}$. Daraus folgt der Anteil der Erwerbstätigen $\mu_{KG_{Kreis}}$ für die Branche KG in dem betrachteten Kreis. Gleichung (2) stellt dies dar:

$$\mu_{KG_{Kreis}} = \frac{n_{KG_{Kreis}}}{n_{Gesamt_{Kreis}}} \quad (2)$$

Um aus dieser Information die Charakteristik einer Branche für einen Kreis abzuleiten, wird das entsprechende bundesweite Mittel, die Bezugsgröße $\mu_{KG_{Bund}}$ bestimmt. Der Quotient $\mu_{KG_{Bund}}$, der durch Division der Anzahl der bundesweit in der entsprechenden Branche KG Erwerbstätigen $n_{KG_{Bund}}$ und der Gesamtzahl der bundesweit Erwerbstätigen $n_{Gesamt_{Bund}}$ berechnet wird, beschreibt den Anteil der Erwerbstätigen, bezogen auf alle Kreise Deutschlands. Gleichung (3) stellt dies dar:

$$\mu_{KG_{Bund}} = \frac{n_{KG_{Bund}}}{n_{Gesamt_{Bund}}} \quad (3)$$

Um zu bestimmen, inwiefern eine Branche KG für einen Kreis charakteristisch ist, wird die Streuung $\sigma_{KG_{Kreis}}$ berechnet:

$$\sigma_{KG_{Kreis}} = \mu_{KG_{Kreis}} - \mu_{KG_{Bund}} \quad (4)$$

Wir interpretieren eine Branche als charakteristisch für einen Kreis, wenn die Streuung einen positiven Wert annimmt, d.h. der Anteil der Erwerbstätigen in diesem Kreis ist überdurchschnittlich hoch. Abb. 3a stellt die diskretisierten Streuungswerte aller Kreise für die Kenngröße „familiäre Beherbergungsbetriebe“ mittels eines Histogramms dar. Der Großteil der Kreise hat eine negative Streuung, d.h. die entsprechende Branche ist in diesen Kreisen nur unterdurchschnittlich vertreten und wird somit als nicht charakteristisch für diese Kreise interpretiert. Das Histogramm von Abb. 3a zeigt eine typische Verteilung der Streuung für eine Kenngröße. Um diese Informationen im weiteren Verlauf nutzen zu können, werden alle Kreise für jede Kenngröße entsprechend der Streuung in fünf Kategorien eingeteilt. Hierfür werden diejenigen Streuungswerte σ_{25} , σ_{50} und σ_{75} berechnet, die der Streuung des $0,25 \cdot n_{\sigma>0}$ -ten, $0,5 \cdot n_{\sigma>0}$ -ten und $0,75 \cdot n_{\sigma>0}$ -ten Kreises entsprechen, wobei es sich bei dem Faktor $n_{\sigma>0}$ um die Anzahl der Kreise mit positiver Streuung handelt. Damit werden bei der Zuweisung der Kategorien alle Kreise miteinander in Bezug gesetzt. Ergebnis ist folgende Einteilung in Kategorien für eine beliebige Kenngröße KG :

$$\begin{aligned} kat0 & \text{ für } \sigma_{KG_{Kreis}} < 0 \\ kat1 & \text{ für } 0 \leq \sigma_{KG_{Kreis}} < \sigma_{25} \\ kat2 & \text{ für } \sigma_{25} \leq \sigma_{KG_{Kreis}} < \sigma_{50} \\ kat3 & \text{ für } \sigma_{50} \leq \sigma_{KG_{Kreis}} < \sigma_{75} \\ kat4 & \text{ für } \sigma_{KG_{Kreis}} \geq \sigma_{75} \end{aligned} \quad (5)$$

Mittels der Kategorien wird der Zustandsraum einer Kenngrößen-Zustandsvariable definiert:

$$x_{KG} = \begin{cases} kat0 & \text{für } kat0 \\ kat1_4 & \text{für } kat1, kat2, kat3, kat4 \end{cases} \quad (6)$$

Abbildung 2b ist eine Visualisierung der Kenngröße „familiäre Beherbergungsbetriebe“ für alle Kreise Deutschlands. Die weiß dargestellten Kreise entsprechen der Kategorie *kat0*. Je dunkler die Darstellung eines Kreises, desto höher die Kategorie, d.h. umso charakteristischer ist die betrachtete Branche für den entsprechenden Kreis.

Die Modelle zu Berechnung der Korrelation zwischen Kenngrößen und Situationsparameter werden als *Kenngrößen-Modelle* bezeichnet.

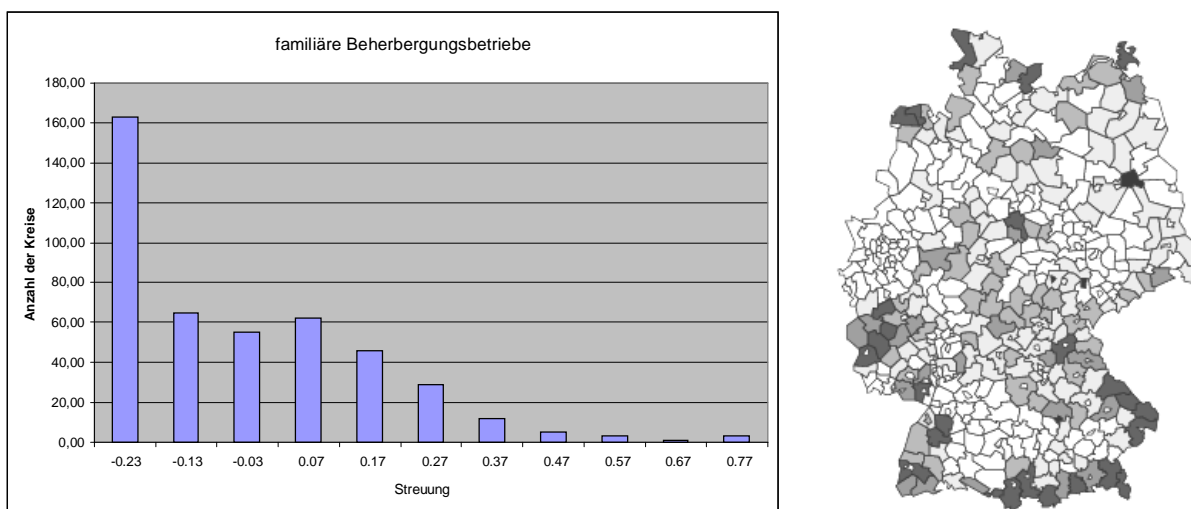


Abb. 2: Klassifikation aller Kreise in Kategorien der Kenngröße *familiäre Beherbergungsbetriebe*

a) Histogramm der Streuungen aller Kreise; b) Visualisierung der Kategorie aller Kreise

Fig. 2: Classification of all areas in categories of the parameter *private accommodation*

a) Histogram of the dispersions of all areas, b) Visualization of the category of all areas

3.4.2. Nutzerspezifische Informationen

Neben den ortsspezifischen Informationen sollen auch noch verschiedene nutzerspezifische Informationen zur Zielort-Prädiktion herangezogen werden. Der Aktionsradius des Fahrers wird mit Hilfe des *Distanz-Modells* situationsabhängig modelliert. Hintergrund ist die Vermutung einer Korrelation zwischen Situationen und durchschnittlicher Entfernung der Zielorte.

Die zweite nutzerspezifische Information ist der „virtuelle“, örtliche Schwerpunkt des Fahrerinteresses. Grundlage hierfür ist ein Koordinatensystem, das auf Kreisebene die Positionen aller Orte beschreibt. Der Schwerpunkt entspricht dabei dem Ort, der die kürzeste Distanz zu allen bereits angefahrenen Orten besitzt. Dieser Ort kann als Mittelpunkt eines Polygons, der durch die Koordinaten aller bereits beobachteten Zielorte

beschrieben wird, interpretiert werden. Mit jedem neu eingegebenen Zielort wird der Schwerpunkt mittels entsprechend gewichteter Vektoraddition auf den neuen Zielort angepasst.

Zur Auflösung der in Abschnitt 2 beschriebenen paradoxen Situation wird anhand des *Default-Modells* das in der aktuellen Situation wahrscheinlichste Ziel bestimmt. Vor allem für regelmäßig angefahrne Zielorte ist diese Systemfunktion interessant.

Auf die zum Einsatz kommenden verschiedenen Modell-Typen wird in den folgenden Abschnitten detailliert Bezug genommen.

3.5 Bayes'sche Netze als Grundlage der Korrelationsmodelle

3.5.1 Topologie der Bayes'schen Netze

Nach der Definition des Situationsbegriffs und der auswertbaren Informationen wird nun der eigentliche Kern des Verfahrens, die Korrelationsmodelle, erläutert. Die mathematische Grundlage für die Modellierung der Korrelation zwischen Situation und Zielort-Informationen bilden Bayes'sche Netze, um potenzielle Korrelationen quantitativ erfassen zu können, da in der Regel kein deterministischer Zusammenhang zwischen Situation und Zielort-Information besteht und somit die Fähigkeit des Systems, mit Unsicherheit umzugehen, gewährleistet sein muss. Zudem muss das Verfahren in der Lage sein, unvollständige Situationsvektoren handhaben zu können.

Bei den Korrelationsmodellen wird zwischen drei Modell-Typen unterschieden: Die Kenngrößen-Modelle, das Distanz-Modell und das Default-Modell.

Zunächst wird die Topologie der Kenngrößen-Modelle detailliert beschrieben. Um einen Zusammenhang zwischen Situation und Zielort herzustellen, wird die Korrelation von der aktuellen Situation und aller den entsprechenden Zielort beschreibenden Kenngrößen ermittelt, d.h. für jede Kenngröße ist ein Modell nötig. Folglich werden für die 28 Kenngrößen entsprechend viele Bayes'sche Netze eingesetzt. Um Speicher- und Rechenaufwand für den Trainings- sowie für den Inferenz-Modus im Rahmen zu halten, wurden die Strukturen der Netzwerke möglichst einfach gehalten. Abb. 3 zeigt die Struktur eines Bayes'schen Netzes zur Modellierung der Korrelation einer Kenngröße und den einzelnen Situationsparametern.

Bei einem Kenngrößen-Modell besteht der Zustandsraum des Wurzelknotens aus den möglichen Kategorien der entsprechenden Kenngröße, also den in Gleichung (6) erwähnten Zuständen *kat0* und *kat1_4*. Die a-priori-Wahrscheinlichkeitsverteilung des Wurzelknotens wird neutral gewählt. Bei den Zustandsvariablen der Situationsparameter handelt es sich um

Bool'sche Variablen, deren bedingte Wahrscheinlichkeiten $P(x_{Sit} | x_{KG})$ zu Beginn neutral definiert werden, im Rahmen des Trainingsmodus schließlich aber auf jeden neuen Datensatz adaptiert werden.

Die Situationsparameter werden mittels Bool'scher Zustandsvariablen repräsentiert. Soll ein Situationsaspekt auf eine Zustandsvariable abgebildet werden, so wird diese Variable auf den Zustand "ja" gesetzt, d.h. im Sommer wird die Zustandsvariable "Sommer" gleich "ja" gesetzt, die Zustandsvariablen der anderen Jahreszeiten bleiben unberücksichtigt; sie werden so behandelt, als liege keine Information bezüglich dieser Situationsparameter vor. Die Struktur spiegelt die Philosophie des Assistenzsystems wider, den Zusammenhang jedes einzelnen Situationsparameters mit einer Kenngröße zu modellieren. Dadurch erhält das System die Fähigkeit, über bisher nicht exakt beobachtete Situationen zu urteilen. Dieser Aspekt ist von erheblicher Bedeutung, da somit der "Einschwingvorgang" des Systems maßgeblich reduziert wird, da nicht exakt jede Situation beobachtet werden muss, bevor die bedingten Wahrscheinlichkeiten $P(x_{Sit} | x_{KG})$ aller Situationsparameter gegen repräsentative Werte konvergieren. Der Einfluss einer Situation wird also auf alle beobachteten Situationsparameter gleichmäßig verteilt.

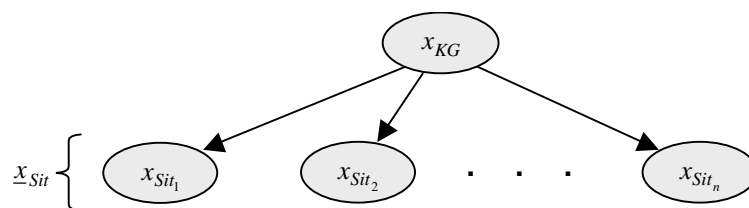


Abb. 3: Topologie des Bayes'schen Netzes zur Modellierung der Korrelation zwischen Situationsparameter und Orts-Kenngröße

Fig. 3: Topology of the Bayesian network for modelling the correlation between situation parameter and location parameter

Die Strukturen der anderen Korrelationsmodelle werden analog gewählt. Der Zustandsraum des Wurzelknotens des Distanzmodells umfasst in *Adaptive Compass* einen Zustand pro Kreisring, für das Default-Modell wird jeder potentielle Default-Ort durch einen Zustand repräsentiert. Die a-priori-Wahrscheinlichkeitsverteilungen der Wurzelknoten aller Modelle werden neutral gewählt.

Im nächsten Abschnitt wird erläutert, wie die bedingten Wahrscheinlichkeiten adaptiv berechnet werden.

3.5.2 Training der Bayes'schen Netze

Der vorhergehende Abschnitt stellte die Struktur der Netzwerke vor, nun wird auf die entsprechenden bedingten Wahrscheinlichkeiten näher eingegangen. Abb. 4 stellt hierfür schematisch das Online-Training der Korrelations-Modelle bzw. der Bayes'schen Netze dar.

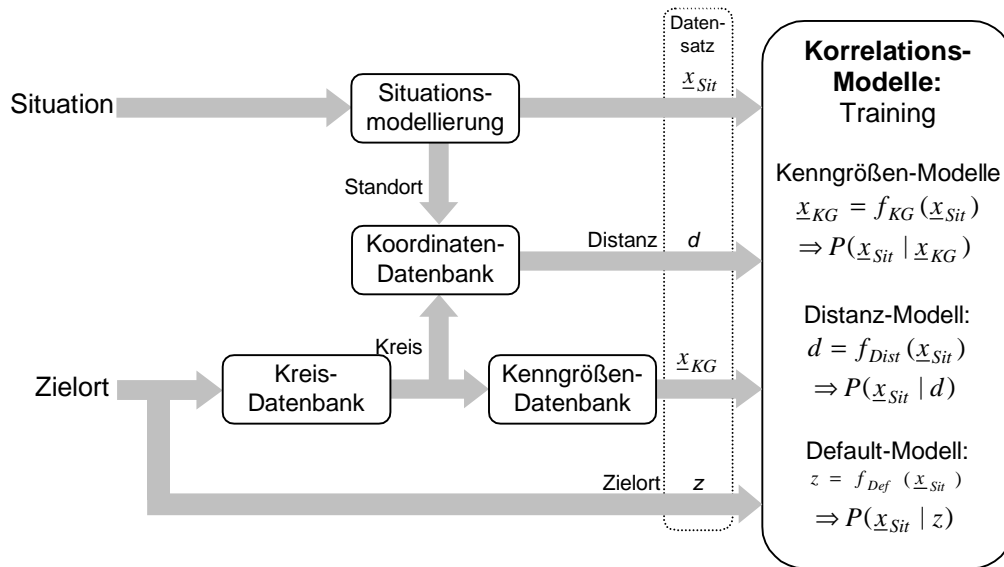


Abb. 4: Training der Korrelationsmodelle

Fig. 4: Training of the correlation models

Gibt der Fahrer in einer bestimmten Situation einen Zielort ein, so interpretiert das System die Beschreibung dieser Situation \underline{x}_{Sit} , die Charakterisierung des eingegebenen Ortes \underline{x}_{KG} sowie die Entfernung dieses Zielorts zur aktuellen Position d und den Zielort z als neuen Datensatz. Die Aufgabe des Trainingsverfahrens besteht in der iterativen Adaption der bedingten Wahrscheinlichkeiten der Situationsparameter $P(x_{Sit} | x_{KG})$ für die Kenngrößen-Modelle, $P(x_{Sit} | d)$ für das Distanzmodell und $P(x_{Sit} | z)$ für das Default-Modell an diesen neuen Datensatz, wobei den vorherigen Datensätzen Rechnung zu tragen ist.

Der Trainingsalgorithmus wurde auf die verwendeten Netzwerkstrukturen und auf die Modellierung der Korrelation zwischen Situation und Zielort zugeschnitten. Das Training eines Kenngrößen-Modells ist in Abb. 5 schematisch dargestellt und wird nun in den Schritten ① bis ⑤ beschrieben:

① Zunächst wird die a-posteriori-Wahrscheinlichkeit $P_{n+1}(x_{KG} | \underline{x}_{Sit})$ berechnet, die sowohl dem neuen Datensatz, als auch vorherigen Datensätzen Rechnung trägt. Dies geschieht durch folgende Gleichung:

$$P_{n+1}(x_{KG} | \underline{x}_{Sit}) = \frac{P_{Data}(x_{KG}) + w_{Data} \cdot P_n(x_{KG} | \underline{x}_{Sit})}{1 + w_{Data}} \quad (7)$$

Die Indizes n und $n+1$ sollen verdeutlichen, dass die betrachteten Wahrscheinlichkeiten bereits beobachtete Datensätze (n), bzw. auch den neuen Datensatz ($n+1$) repräsentieren. $P_{Data}(x_{KG})$ ist direkt dem neuen Datensatz zu entnehmen und entspricht, je nach Zustand, entweder dem Wert 0 oder 1. Mit dem Faktor w_{Data} wird der Einfluss des neuen Datensatzes im Verhältnis zu bisherigen Beobachtungen festgelegt. Orte mit hohen Kategorien bezüglich einer Kenngröße werden für das Training durch einen niedrigeren Faktor w_{Data} im Verhältnis zu vorherigen Datensätzen stärker gewichtet.

Ziel des Trainings-Verfahrens ist, die bedingten Wahrscheinlichkeiten der Situationsparameter $P_n(x_{Sit} | x_{KG})$ iterativ so zu verändern, dass die für den Inferenzmodus interessante Wahrscheinlichkeit $P_n(x_{KG} | \underline{x}_{Sit})$ der Wahrscheinlichkeit $P_{n+1}(x_{KG} | \underline{x}_{Sit})$ entspricht.

② Der mit den aktuellen Wahrscheinlichkeiten $P_n(x_{Sit} | x_{KG})$ korrespondierende Wert $P_n(x_{KG} | \underline{x}_{Sit})$ wird berechnet.

③ Schließlich wird die Abweichung ΔP von Soll- und Ist-Zustand der a-posteriori-Wahrscheinlichkeit für die Kategorie kat der betrachteten Kenngröße bestimmt:

$$\Delta P = \left| P_{n+1}(x_{KG} = kat | \underline{x}_{Sit}) - P_n(x_{KG} = kat | \underline{x}_{Sit}) \right| \quad (8)$$

④ Sobald alle Werte der Differenz ΔP eine Toleranzgrenze D_{konv} unterschreiten, deutet dies auf eine Konvergenz von $P_n(x_{KG} | \underline{x}_{Sit})$ gegen ihren Soll-Wert $P_{n+1}(x_{KG} | \underline{x}_{Sit})$ hin. Das Abbruchkriterium der Iterationsschleife ist erfüllt und das Netzwerk approximativ auf den neuen Datensatz trainiert. Bei der Wahl der Konstanten D_{konv} ist dabei ein Kompromiss zwischen Genauigkeit und niedriger Anzahl von Iterationsschritten zu treffen.

Ist das Abbruchkriterium der Iterationsschleife nicht erfüllt, so werden die Wahrscheinlichkeitswerte in nächsten Schritt verändert.

⑤ Gleichung (9) erlaubt eine Adaption der Wahrscheinlichkeitswerte $P_n(x_{Sit} | x_{KG})$ in Analogie zu Gleichung (7):

$$P_{n+1}(x_{Sit} | x_{KG}) = \frac{P_{Data}(x_{Sit}) + w_{Sit} \cdot P_n(x_{Sit} | x_{KG})}{1 + w_{Sit}} \quad (9)$$

Möglich sind allerdings nur Aussagen über die Werte der bedingten Wahrscheinlichkeit $P_n(x_{Sit} | x_{KG})$, die mit den Zuständen der Zustandsvariablen aus dem neuen Datensatz korrespondieren, d.h. bei Kenngrößen-Netzwerken werden nur die Werte der beobachteten Kategorie der Kenngröße adaptiert, weil durch die Netzwerk-Topologie die möglichen Zustände einer Kenngröße, abhängig von der Situation, nicht in Rivalität treten. Um dies zu ermöglichen, werden zusätzliche, "künstliche" Datensätze erzeugt, um quantitative Aussagen über alle Einträge der Wahrscheinlichkeitstabelle machen zu können. Für jeden

im aktuellen Datensatz nicht beobachteten Zustand des Wurzelknotens eines Netzwerks wird ein Datensatz geschaffen, der den entsprechenden Zustand aktiviert und außerdem die Situationsvariable auf den Zustand "nein" setzt. Somit kann ein Situationsparameter für einen bestimmten Zustand des Wurzelknotens mehr und für die übrigen Zustände entsprechend weniger charakteristisch sein.

Die Werte werden, ausgehend von ihrem ursprünglichen Wert, durch den Faktor w_{Sit} in Richtung des für die Konvergenz erforderlichen Wertes entwickelt, indem Gleichung (9) auf alle neuen Datensätze angewendet wird.

Ausgehend für von einem Startwert 0 für w_{Sit} wird dieser Faktor in jedem Iterationsschritt erhöht. Auf Optimierungsmaßnahmen mittels dynamischer Anpassung der Schrittweite wird in dieser Veröffentlichung nicht näher eingegangen.

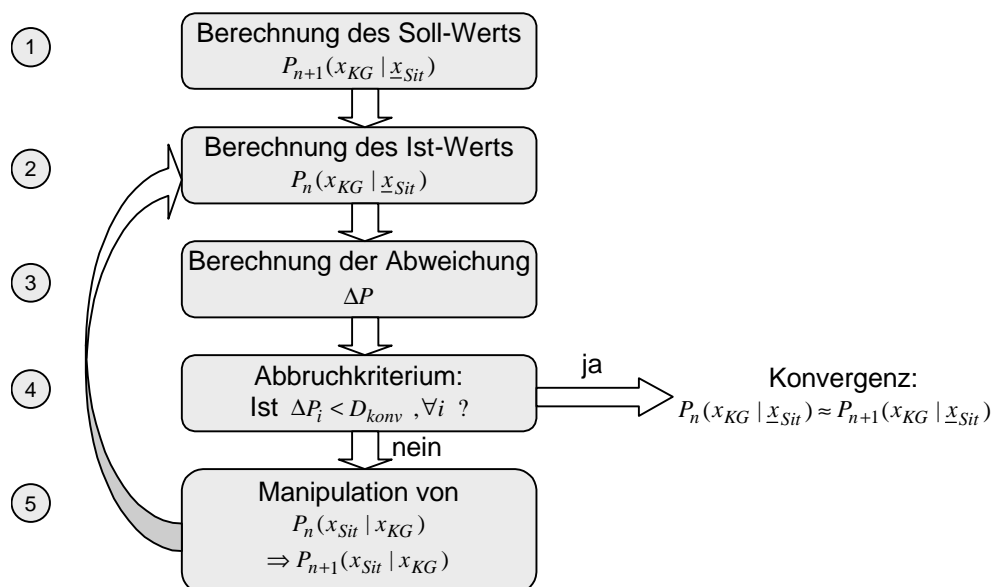


Abb. 5: Iteratives Training der Bayes'schen Netze

Fig. 5: Iterative training of the Bayesian networks

Das Training des Distanz-Modells und des Default-Modells wird analog zu dem oben beschriebenen Verfahren durchgeführt.

3.6 Zielort-Prädiktion

3.6.1 Evaluierung aller Orte/Kreise

Für die situative Zielort-Prädiktion erfolgt eine Abbildung der aktuellen Situation auf den Zustandsvektor x_{Sit} , der schließlich als Eingangsvektor für die Korrelationsmodelle dient. Im Inferenzmodus werden alle Beobachtungen bezüglich der Situation auf die entsprechenden

Korrelationsmodelle abgebildet und die interessierenden Wahrscheinlichkeitsverteilungen berechnet. Abb. 6 zeigt das Prinzip der situativen Evaluierung aller Kreise Deutschlands.

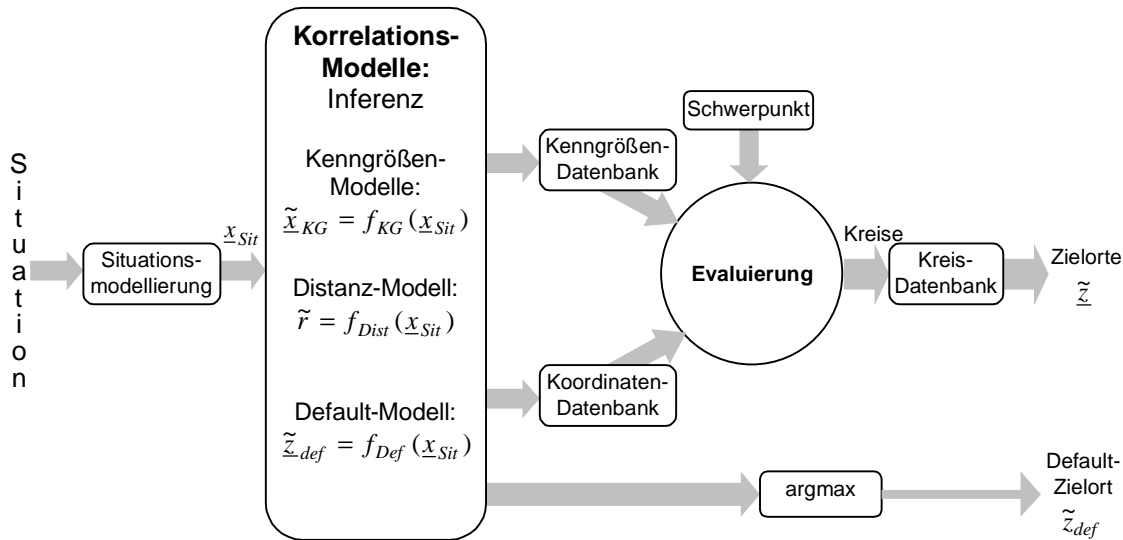


Abb. 6: Situative Evaluierung aller Kreise Deutschland

Fig. 6: Situative evaluation of all areas of Germany

Aus den Korrelations-Modellen und aus den anderen fahrerspezifischen Informationsquellen werden nun Evaluierungsfaktoren berechnet. Aus den Kenngrößen-Modellen wird E_{KG} bestimmt, aus dem Distanz-Modell E_{Radius} ; nutzerspezifische Informationen werden in dem Faktor E_{Fahrer} zusammengefasst. Durch Multiplikation dieser Evaluierungsfaktoren erhält man das Evaluierungsergebnis eines Kreises E_{Kreis} :

$$E_{Kreis} = E_{KG} \cdot E_{Radius} \cdot E_{Fahrer} \quad (10)$$

Nun wird die Berechnung der einzelnen Evaluierungsfaktoren E_{KG} , E_{Radius} und E_{Fahrer} im Detail beschrieben.

Für die Bestimmung von E_{KG} werden die auf Basis bisheriger Beobachtungen wahrscheinlichsten Kategorien \tilde{x}_{KG} für alle 28 Kenngrößen berechnet. Diese Informationen und die Kenngrößen-Datenbank werden schließlich zur Evaluierung aller Kreise verwendet. Hierfür wird zunächst die quantitative Annahme $Bel_{Kreis}(x_{KG})$ berechnet, dass der zu evaluierende Kreis für den Fahrer in der Situation x_{Sit} aufgrund bestimmter charakteristischer Eigenschaften des Kreises von Interesse ist.

Zunächst wird die Annahme $Bel_{Kreis}(x_{KG_i})$ bezüglich einer Kenngröße KG_i berechnet: Hierfür wird die Wahrscheinlichkeit $P(x_{KG_i} = kat1_4 | x_{Sit})$ berechnet und interpretiert. Ein Wert größer als 0,5 modelliert die Bevorzugung von Gegenden mit einem der Kenngröße

entsprechenden überdurchschnittlich stark vertretenen Wirtschaftszweig oder Freizeitaspekt. Diese Wahrscheinlichkeit $P(x_{KG_i} = kat1_4 | \underline{x}_{Sit})$ entspricht dann $Bel_{Kreis}(x_{KG})$. Unterschreitet $P(x_{KG_i} = kat1_4 | \underline{x}_{Sit})$ den Wert 0,5, so lässt sich dies als Gleichgültigkeit gegenüber dem betreffenden Wirtschaftszweig interpretieren; aus diesem Grund wird $Bel_{Kreis}(x_{KG})$ mit dem Wert 0,5 quantitativ beschrieben. Die gleiche Wertzuweisung erfolgt im Falle nicht verfügbarer Informationen bezüglich einer Kenngröße eines bestimmten Kreises. Gleichung (11) fasst die Bestimmung von $Bel_{Kreis}(x_{KG})$ bezüglich der Kenngröße KG_i zusammen:

$$Bel_{Kreis}(x_{KG_i}) = \begin{cases} P(x_{KG_i} | \underline{x}_{Sit}) & \text{für } P(x_{KG_i} = kat1_4 | \underline{x}_{Sit}) \geq 0,5 \\ 0,5 & \text{für } P(x_{KG_i} = kat1_4 | \underline{x}_{Sit}) < 0,5 \\ 0,5 & \text{falls Kategorie unbekannt} \end{cases} \quad (11)$$

Diese Evaluierung wird für alle 28 Kenngrößen durchgeführt. Unter der Annahme statistischer Unabhängigkeit der Kenngrößen führt dies zu folgender Gleichung:

$$E_{KG} = \prod_{i=1}^{28} Bel_{Kreis}(x_{KG_i}) \quad (12)$$

Für den Evaluierungsfaktor E_{Radius} wird mit Hilfe des Distanz-Modells der fahrertypische Aktionsradius situativ modelliert. Zunächst wird die Wahrscheinlichkeitsverteilung $P(r | \underline{x}_{Sit})$ berechnet, um Aufschluss über die in entsprechenden Situationen typischerweise zurückgelegten Distanzen zu erhalten. Mittels Gleichung (13) wird durch das Maß $Bel(r_{Aktion})$ eine quantitative Aussage über den geschätzten Aktionsradius in der aktuellen Situation möglich. Hierfür werden die Wahrscheinlichkeiten der einzelnen Kreisringe, abhängig von der Entfernung des zu evaluierenden Kreises vom aktuellen Standort, in folgender Weise addiert:

$$E_{Radius} = Bel(r_{Radius}) = \begin{cases} 1 & d_{Kreis} < 50km \\ P(d = d_{50-100} | \underline{x}_{Sit}) + P(d = d_{100-250} | \underline{x}_{Sit}) \\ \quad + P(d = d_{250-500} | \underline{x}_{Sit}) + P(d = d_{\geq 500} | \underline{x}_{Sit}) & 50km \leq d_{Kreis} < 100km \\ P(d = d_{100-250} | \underline{x}_{Sit}) + P(d = d_{250-500} | \underline{x}_{Sit}) \\ \quad + P(d = d_{\geq 500} | \underline{x}_{Sit}) & 100km \leq d_{Kreis} < 250km \quad (13) \\ P(d = d_{250-500} | \underline{x}_{Sit}) + P(d = d_{\geq 500} | \underline{x}_{Sit}) & 250km \leq d_{Kreis} < 500km \\ P(d = d_{\geq 500} | \underline{x}_{Sit}) & d_{Kreis} \geq 500km \end{cases}$$

Die Indizes der Zustände d_x geben die entsprechenden Distanzen der Kreisringe in Kilometern an.

Die Berücksichtigung offensichtlicher, nicht situativer Interessen des Fahrers erfolgt durch den Evaluierungsfaktor E_{Fahrer} , der dem Aktions-Schwerpunkt des Fahrers, sowie weiteren Informationen Rechnung trägt. Die Koordinate des Aktions-Schwerpunkts gibt den Mittelpunkt des Gebiets an, in dem sich der Fahrer im allgemeinen aufhält. Orte, deren Entfernung s_{Kreis} zu diesem Schwerpunkt gering ist, werden in der Evaluierung daher stärker gewichtet als weiter entfernte Orte. Gleichung (14) beschreibt die Berechnung des verwendeten Evaluierungsfaktors E_{Fahrer} :

$$E_{Fahrer} = \left(1 - \frac{s_{Kreis}}{2s_{max}} \right) \quad (14)$$

Die Entfernung s_{max} entspricht dabei der größten beobachteten Entfernung unter allen Kreisen, um die Evaluierung eines Kreises mit den anderen Kreisen in Bezug zu setzen. Durch den Faktor E_{Fahrer} werden noch weitere Informationen aus dem Umfeld des Fahrers modelliert.

Zusammenfassend wird durch die Gleichungen (10), (12), (13) und (14) die Evaluierung eines Kreises folgendermaßen vorgenommen:

$$E_{Kreis} = \prod_{i=1}^{28} Bel_{Kreis}(x_{KG_i}) \cdot Bel(r_{Aktion}) \cdot \left(1 - \frac{s_{Kreis}}{2s_{max}} \right) \quad (15)$$

Schließlich werden die Evaluierungsmaße aller Kreise noch auf 1 normiert.

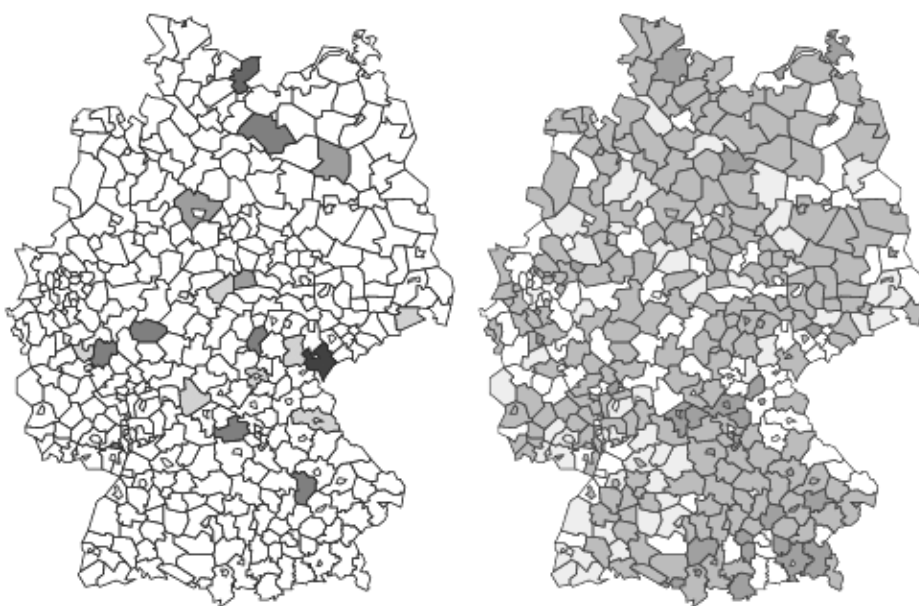


Abb. 7: a) Disambiguierung mehrdeutiger Eingaben; b) Evaluierung aller Kreise zur Vokabularbegrenzung
 Fig. 7: a) Disambiguation of ambiguous inputs; b) Evaluation of all areas for limiting the location-vocabulary

Für die Disambiguierung werden alle Alternativen evaluiert und der wahrscheinlichste Zielort dem Fahrer angeboten. Abb. 7a zeigt die Evaluierungsergebnisse der Kreise, die ein „Neustadt“ beinhalten. Je dunkler die Darstellung, desto größer die Annahme, dass es sich um das korrekte „Neustadt“ handelt. In dem dargestellten Fall wurden „Neustadt bei Holstein“ und „Neustadt im Vogtland“ mit ähnlichem Ergebnis evaluiert. Somit werden dem Fahrer diese beiden Städte zur Auswahl angeboten.

Das Ergebnis der Evaluierung kann nun zur Vokabularbegrenzung und zur Nachbearbeitung der Spracherkenner-Ergebnisse verwendet werden. Abb. 7b zeigt die entsprechenden Evaluierungsergebnisse.

3.6.2 Prädiktion des Default-Zielorts

Zur Berechnung des Default-Zielorts wird die aktuelle Situation auf das Default-Modell abgebildet. Jeder Zustand des Wurzelknotens entspricht einer Default-Zielort-Hypothese, d.h. jeder Eintrag der Zielort-History wird durch einen Zustand repräsentiert. Wir interpretieren den Ort mit der größten a-posteriori-Wahrscheinlichkeit als den Default-Zielort \tilde{z}_{def} . Gleichung (16) stellt die mathematisch dar:

$$\tilde{z}_{def} = \arg \max_i \{P(z_i | \underline{x}_{Sit})\} \quad (16)$$

Dieses Ergebnis wird dem Fahrer angeboten, sobald er in das Navigationsmenu wechselt und der zuletzt eingegebene Zielort dem aktuellen Standort entspricht. Im Falle einer korrekten Prädiktion braucht der Fahrer den Zielort lediglich zu bestätigen; hat der Fahrer ein anderes Ziel, so kann der Default-Zielort ignoriert werden und das tatsächliche Ziel eingegeben werden.

3.7 Ergebnisse und Diskussion

Um den Navigationsassistenten *Adaptive Compass* adäquat zu testen und zu evaluieren ist die Protokollierung der Fahrgewohnheiten in Abhängigkeit von allen Situationsparametern über einen möglichst großen Zeitraum nötig. Zudem wäre die Analyse von Personen mit unterschiedlichstem beruflichen und privaten Hintergrund sinnvoll. Die Akquirierung dieser Daten bedeutet immensen Aufwand und künstlich erzeugte Datensätze können kaum authentische Verhältnisse widerspiegeln.

Da gegenwärtig keine Daten aus größeren Versuchsreihen zur Verfügung stehen, konnte *Adaptive Compass* nur auf Basis der Daten einer Person, des Entwicklers des Assistenzsystems, evaluiert werden. Evaluiert wurden die Disambiguierung von Ortsnamen sowie die Default-Zielort-Prädiktion.

Im eingeschwungenen Zustand, d.h. im Fall sinnvoll konvergierter bedingter Wahrscheinlichkeiten der Korrelationsmodelle konnten die Ambiguitäten zu 75 Prozent auf Anhieb aufgelöst werden. In diesen Fällen wurde eine Zielort-Alternative erheblich besser evaluiert als die übrigen und folglich dem Fahrer direkt, mit der Option der Ablehnung, angeboten. Wird eine Reihe von Alternativen besser evaluiert als die übrigen, so wird diese Liste dem Fahrer zu Auswahl angeboten. Die Wahrscheinlichkeit, dass das tatsächliche Wunschziel Teil dieser Liste war, betrug 90 Prozent. Diese situationsadaptiv erstellte Liste enthielt in allen Testsituationen um ein Vielfaches weniger Ziel-Alternativen als die ursprüngliche, komplette Liste von Zielorten. Zusammenfassend lässt sich feststellen, dass der Navigationsassistent in der Lage ist, in 90 Prozent der Situationen, in denen der Fahrer einen mehrdeutigen Zielort eingegeben hat, einen erheblich kürzeren, schnelleren, komfortableren und sicheren Dialog mit dem Navigationssystem zu ermöglichen.

Der Default-Zielort konnte mit einer Quote von 40 Prozent richtig erkannt und dem Fahrer angeboten werden. Zu berücksichtigen ist dabei, dass es sich bei der Testperson um keinen Pendler handelt, der regelmäßig, d.h. mehrmals pro Woche zwischen zwei Orten pendelt, sondern eher unregelmäßiges Verhalten an den Tag legt. Somit ist eine situative Prädiktion mit 40 Prozent Erfolgsquote als außerordentlich positives Ergebnis zu werten.

Die Evaluierung aller Kreise zur Erhöhung der Erkennungsrate des Spracherkenners wurde nicht explizit getestet, da das Verfahren dem Algorithmus zur Disambiguierung von Ortsnamen gleicht.

4. Ausblick

Das entwickelte Expertensystem ist in Navigationssystemen der Zukunft als zentrale Intelligenz denkbar. Nicht nur die Kommunikation wird erheblich komfortabler und sicherer, das Wissen des Expertensystems kann auch zur fahrer- und situationsadaptiven Routenplanung verwendet werden. Der nächste Schritt ist eine intensive Datenakquirierung für eine ausführlichere Evaluierung von *Adaptive Compass*.