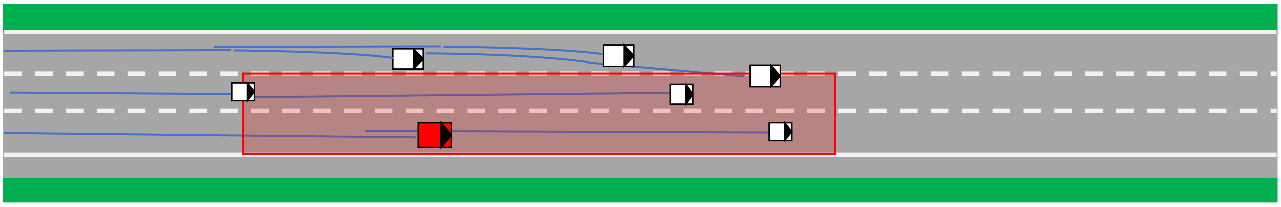
Improvement and validation of a highway traffic complexity metric for test scenarios of automated vehicles



Scientific work for obtaining the academic degree

Master of Science (M.Sc.)

at the Department of Mechanical Engineering of the Technical University of Munich

Supervised by Prof. Dr.-Ing. Markus Lienkamp

Thomas Ponn, M.Sc.

Chair of Automotive Technology

Submitted by Xiao Yu, B.Sc.

Boltzmannstraße 15

85748 Garching bei München

Submitted on 16.06.2020

Lehrstuhl für Fahrzeugtechnik
Fakultät für Maschinenwesen
Technische Universität München

Description of the work

A metric being able to evaluate and quantify the complexity of traffic situation helps to shorten the process of selecting the most compact and representative test cases for automated vehicles. The developed metric in the existing work calculates an indicator for complexity by linear combination of influence of different factors. However, there are several shortcomings in this metric, which make the results of the metric less convincing. To improve these shortcomings will be the focus of this work

The following tasks are to be completed by Ms. Xiao Yu:

- Improvement of some shortcomings in the previous work
- Extensions of the previous work with several more influence factors
- Sensitivity analysis of the metric
- Determination of weighting factors of different influence factors

Each step of the work should be documented in a clear form. The candidate undertakes to finish the master thesis independently and to indicate the scientific tools used by her.

The submitted work remains as the examination document in the ownership of the chair.

Release: 02.12.2019 Submit: 16.06.2020

Prof. Dr.-Ing. M. Lienkamp Supervisor: Thomas Ponn, M. Sc.

Lehrstuhl für Fahrzeugtechnik
Fakultät für Maschinenwesen
Technische Universität München

Geheimhaltungsverpflichtung

Frau: **Yu, Xiao**

Gegenstand der Geheimhaltungsverpflichtung sind alle mündlichen, schriftlichen und digitalen Informationen und Materialien, die der Unterzeichner vom Lehrstuhl oder von Dritten im Rahmen seiner Tätigkeit am Lehrstuhl erhält. Dazu zählen vor allem Daten, Simulationswerkzeuge und Programmcode sowie Informationen zu Projekten, Prototypen und Produkten.

Der Unterzeichner verpflichtet sich, alle derartigen Informationen und Unterlagen, die ihm während seiner Tätigkeit am Lehrstuhl für Fahrzeugtechnik zugänglich werden, strikt vertraulich zu behandeln.

Er verpflichtet sich insbesondere:

- derartige Informationen betriebsintern zum Zwecke der Diskussion nur dann zu verwenden, wenn ein ihm erteilter Auftrag dies erfordert,
- keine derartigen Informationen ohne die vorherige schriftliche Zustimmung des Betreuers an Dritte weiterzuleiten,
- ohne Zustimmung eines Mitarbeiters keine Fotografien, Zeichnungen oder sonstige Darstellungen von Prototypen oder technischen Unterlagen hierzu anzufertigen,
- auf Anforderung des Lehrstuhls für Fahrzeugtechnik oder unaufgefordert spätestens bei seinem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik alle Dokumente und Datenträger, die derartige Informationen enthalten, an den Lehrstuhl für Fahrzeugtechnik zurückzugeben.

Besondere Sorgfalt gilt im Umgang mit digitalen Daten:

- Für den Dateiaustausch dürfen keine Dienste verwendet werden, bei denen die Daten über einen Server im Ausland geleitet oder gespeichert werden (Es dürfen nur Dienste des LRZ genutzt werden (Lehrstuhllaufwerke, Sync&Share, GigaMove).
- Vertrauliche Informationen dürfen nur in verschlüsselter Form per E-Mail versendet werden.
- Nachrichten des geschäftlichen E-Mail Kontos, die vertrauliche Informationen enthalten, dürfen nicht an einen externen E-Mail Anbieter weitergeleitet werden.
- Die Kommunikation sollte nach Möglichkeit über die (my)TUM-Mailadresse erfolgen.

Die Verpflichtung zur Geheimhaltung endet nicht mit dem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik, sondern bleibt 5 Jahre nach dem Zeitpunkt des Ausscheidens in vollem Umfang bestehen. Die eingereichte schriftliche Ausarbeitung darf der Unterzeichner nach Bekanntgabe der Note frei veröffentlichen.

Der Unterzeichner willigt ein, dass die Inhalte seiner Studienarbeit in darauf aufbauenden Studienarbeiten und Dissertationen mit der nötigen Kennzeichnung verwendet werden dürfen.

Datum: 02.12.2019

Unterschrift: Yu Xiao

Erklärung

Ich versichere hiermit, dass ich die von mir eingereichte Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Garching, den 16.06.2020

Yu Xiao

Xiao Yu, B. Sc.

Declaration of Consent, Open Source

Hereby I, **Yu, Xiao**, born on **15.09.1992**, make the software I developed during my **master thesis** available to the Institute of Automotive Technology under the terms of the license below.

Garching, 02.12.2019

Yu Xiao

Xiao Yu, B. Sc.

Expression of thanks

It is a pleasure to have this opportunity writing my master thesis at Chair of Automotive Technology. Sincere thanks to my supervisor Thomas Ponn, M. Sc., who has offered a lot of support and useful advices for the accomplishment of this work.

Garching, 02 06 2020

Xiao Yu

Inhaltsverzeichnis

List of Abbreviations

ROI	Region of Interest
SRC	Standardized Regression Coefficient
SND	Sigma-Normalized Derivative
ACC	Adaptive Cruise Control
TCI	Trajectory Criticality Index
fps	frames per second
Pa	Possible actions
Noa	Number of actions

Formelzeichen

Formula symbol	Unit	Description
d_{safe}	m	Safety distance
d_{label}	m	Longitudinal distance between the ego-vehicle and surrounding vehicle “$label$”, with $label \in \{leftfollow, rightfollow, leftpre, middlepre, rightpre\}$
$v_{\mathrm{ego,x}}$	m/s	Longitudinal velocity of ego-vehicle
$v_{i,x}$	m/s	Longitudinal velocity of surrounding vehicle i
$\boldsymbol{v_x}$	m/s	Vector with longitudinal velocities of all surrounding vehicles
f_i	-	Normalized value of factor i
w_i	-	Weighting factor of factor i
$w_{\mathrm{vx,i}}$	-	Weighting factor for longitudinal velocity of surrounding vehicle i
$dynamic_{v,x}$	m/s	Dynamic of surrounding vehicles w.r.t. longitudinal velocity
$variation_{v,x}$	m/s	Variation of dynamical parameters of surrounding vehicles w.r.t. surrounding vehicles
$t_{predict}$	s	Time period for the prediction of trajectories of surrounding vehicles
$time\_gap_{label}$	s	Time-gap between the ego-vehicle and a surrounding vehicle “$label$”, with $label \in \{leftfollow, rightfollow, leftpre, middlepre, rightpre\}$
C_{scene}	-	Complexity of a scene
$C_{scenario}$	-	Complexity of a scenario
$N(\overline{x_i}, \sigma_{x_i})$	-	Distribution of variable x_i with mean $\overline{x_i}$ and standard deviation σ_{x_i}
$S^p_{x_i}$	-	Partial derivative of variable x_i to output
$S^\sigma_{x_i}$	-	Sigma-normalized derivative of variable x_i
β_{x_i}	-	Standard regression coefficient of variable x_i
α	-	Cronbach’s Alpha

V_i	-	Variance of influence factor i
V_t	-	Variance of total score
H	-	Test statistic
H_0	-	Null hypothesis
df	-	Degree of freedom
k	-	Number of items/factors
p	-	Significance level
$\overline{R_\iota}$	-	Mean rank of item i

1 Einleitung

In the first part of this chapter, the backgroud study of this thesis will be introduced, The found issues of the previous study will be discussed and the motivation of this thesis will be mentioned. In the second part, the goal to be achieved will be declared in detail and the structure of the work will be illustrated.

1.1 Relevance of the work

The content of this thesis is based on the work of YU [1], in which a metric is developed to quantify the evaluation of traffic complexity in highway scenarios. This metric starts from the analysis of concept "Complexity" and has derived 10 factors which can influence the complexity of the traffic situation from the perspective of the automated ego-vehicle. The metric defines a Region of Interest (ROI), which is a close area around the ego-vehicle and determined by the longitudinal velocity of ego-vehicle. The calculations of all factors are then restricted in this area. The value of each influence factor will be calculated and normalized into a number between 0 and 1. Finally, with identical weighting factors for all influence factors and linear combination of all normalized values, a number between 0 and 1 can be obtained. This number is then considered as the indicator or measurement for the level of complexity of the traffic situation around the ego-vehicle. For instance, scenarios with a result in interval 0 and 1/3, can be seen to have a low level of complexity, those in interval 1/3 and 2/3 are the ones with a medium level of complexity, scenarios with a result larger than 2/3 are considered to be very complex.

Before launching vehicles with automated functions, a large number of various tests need to be conducted to ensure their performance in the real-world traffic. Generation of test scenarios basically follows two principles. One is based on comprehensiveness, which requires expert knowledge and experience. The other is based on real-world traffic data. Scenarios from real-world traffic data can make the test environment closer to the real world. However, these scenarios can be countless in the real world and it is impossible to test them all. With the developed metric the complexity of all scenarios can be calculated. Scenarios with high level of complexity are assumed to be the most difficult ones and can usually lead to critical situations. These scenarios can be used for testing and the process of selecting test cases can therefore be more objective, more efficient and more economic.

However, this metric to quantify the complexity of traffic situation has some flaws and limitations. Some flaws for instance, when evaluating the factor "number of connections between traffic participants within ROI", the ROI is divided into twelve sectors, only connections between vehicles from two sectors next to each other in lateral and longitudinal direction are considered, While in the reality, vehicles in diagonal direction of the ego-vehicle, namely, vehicles in front of ego or behind ego in adjacent lanes can influence ego-vehicle's behaviour with respect to lane change. These connections in diagonal directions of the ego-vehicle need to be taken into consideration as well. One limitation of the metric is that, all ten derived influence factors are scene-based. A scene is a snapshot of the traffic situation, while a scenario is a temporal sequence of consecutive scenes (A clear distinction between scene and scenario can be found in ULBRICH ET.AL's work [2]). Therefore, factors, which can describe the characteristics of a scenario, are missing.

One major problem of the developed metric is that, the ten derived factors are considered as equally important, which is not realistic. Besides, after normalization, values of different factors lie in different ranges. For some factors, it is common to obtain a normalized value of 0.5, while for some, 0.5 can already be seen as a very large value. Considering all factors as equally important can lead to overestimation or underestimation, which will make the scenarios selected through this metric not so convincing and representative. Therefore, this thesis will serve as an improvement and extension of the previous work, where some shortcomings and problems will be improved and perfected.

1.2 Goal and structure of the work

As explained in the last section, this thesis aims to improve and perfect the previous work. The content of this thesis consists of the following chapters. The metric developed in the previous work will be introduced in the chapter "State of the art". The methods for sensitivity analysis which study the influence of different inputs on the output of a system will be included in this chapter as well. The method of the work is explained in the third chapter, which is made up of three sections. The first section will improve the shortcomings in the previous work. In the second section some factors which are not considered in the previous work, but can also influence the complexity of the scenario, will be included in the metric. The third section tries to figure out the influence of each factor has on scenario complexity and the relative importance of each factor. The former will be achieved with the help of a sensitivity analysis. The latter corresponds to the weighting factor of each influence factor, which will be determined with the help of conducting an online survey. A questionnaire will be designed for the survey and will be introduced in this section. In the fourth chapter the results of the questionnaire for the survey will be analyzed and the weighting factors of each influence factor will be determined based on these results. A sensitivity analysis will be conducted one more time with newly determined weighting factors and the result will be compared with the one using equal weighting factors. The work of the thesis is summarized in the fifth chapter. Further discussions related to the topic of this thesis will be mentioned in the last chapter.

2 State of the art

This chapter consists of two parts. In the first part, the metric developed in the previous work will be introduced. In the scecond part, the definition and several methods of sensitivity analysis will be explained, which can be used to study how influential each factor is on the complexity of the scenario.

2.1 Background of the work

As mentioned in the first chapter, the content of this thesis is based on the work of YU [1]. The developed metric of the previous work is based on the following prerequisites, these prerequisites will be also kept for this thesis:

1. The traffic situation is analysed from the perspective of an automated vehicle. It differs from the situation of considering ego-vehicle as a human driving vehicle. On the one hand, the perception of human drivers can be different even for the same traffic situation, since their capabilities to see, to hear and to response are not the same. In contrast, the perception of an automated vehicle is more objective and consistent. On the other hand, a human driver can evaluate the surrounding environment qualitatively, while an automated vehicle can obtain the movement state of a surrounding vehicle quantitatively with help of different sensors.
2. The metric is developed for the purpose of traffic situation analysis in highway scenarios. Whether the metric can be adapted for evaluation of other types of scenarios remains in discussion and will not be included in this thesis.
3. The objects used for the analysis are restricted to dynamic elements of a scenario, namely the different types of traffic participants and their movements, which corresponds to the fourth layer of the five-layer-model developed by BAGSCHIK ET AL. [3] (Figure 2.1). Factors like infrastructure, weather condition, etc. will not be considered.

Same as the previous work, HighD dataset will continue to be used, which is a naturalistic dataset recorded by a drone at different locations of German highways (Detailed introduction of the dataset can be found in the second chapter of the previous study [1, pp. 14-16] and the official website of the dataset [4]).

2.1.1 Influence factors

The highway section recorded by the drone is about 420 meters long and contains two or three lanes. Arbitrary one of the vehicles appearing in the section can be seen as an automated ego-vehicle. The complexity of the scenario which ego-vehicle is in, will be decided by the traffic participants around it. Vehicles with different distances to the ego-vehicle have different degrees of influence on ego-vehicle. Vehicles too far away from ego-vehicle will barely have any influence. Therefore, a Region of Interest (ROI) is defined, which is a close area around ego-vehicle. Only surrounding vehicles appearing in ROI will be taken into consideration. The length of this area is decided by the longitudinal velocity of ego-vehicle, which is twice the safety distance in front of the ego-vehicle and once the safety distance behind the ego-vehicle (Figure 2.2). Safety distance is calculated through the following formula:

$$d_{safe} = 1.8s \cdot v_{ego,x} \tag{2.1}$$

$v_{ego,x}$ is the longitudinal velocity of ego-vehicle in m/s. A graphical representation of ROI is shown in Figure 2.2, which is the area marked with colour red. Its width includes the lane which ego-vehicle is currently in, and the lane left and/or right to ego-vehicle if available. If ego-vehicle is in the middle lane of a three-lane highway, ROI will have a width of three lanes, for other cases (ego-vehicle not in the middle lane of a three-lane highway or ego-vehicle on a two-lane highway) ROI has a width of two lanes.

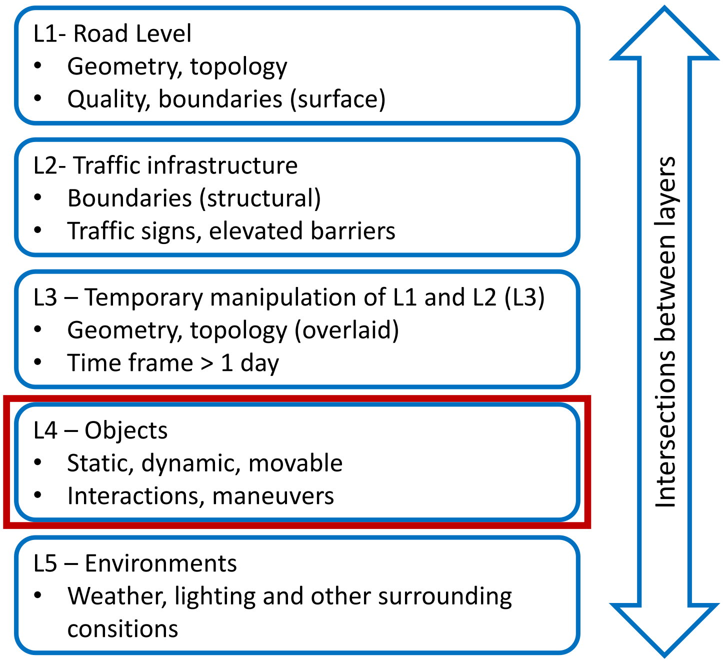


Figure 2.1: Five-layer-Model [3]

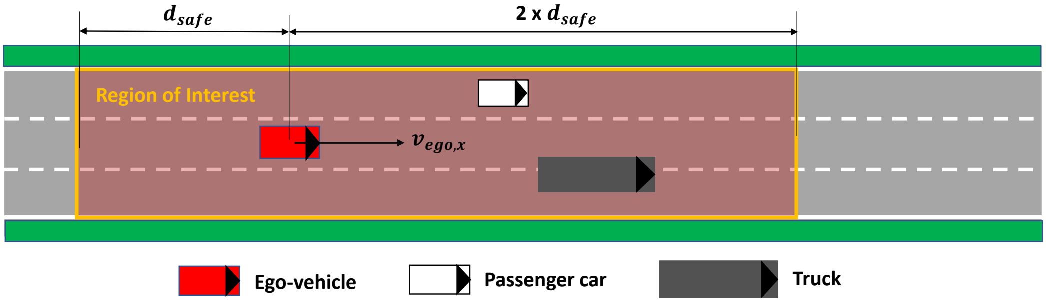


Figure 2.2: Region of Interest

Based on the definition of ROI, only vehicles within ROI will be considered for the evaluation of derived factors, which will be explained in detail in the following text. All factors in the previous work are scene-based, which means the values of these factors can be calculated for every scene of the scenario.

Factor 1: Types of surrounding vehicles

Different types of vehicles have different characteristics of movement, with more types more dynamic can be brought to the scenario. Since in HighD dataset only information of passenger car and truck is available, therefore only these two types will be considered. Thus, total number of types of surrounding vehicles is one of the following three possibilities: 0 (no vehicle around ego-vehicle within ROI), 1 (surrounding vehicles are all passenger cars or all trucks) and 2 (both passenger cars and trucks are present, which is the example shown in Figure 2.2)

Factor 2: Number of surrounding vehicles

This statement is self-explanatory, which is the total number of traffic participants within ROI except ego-vehicle. In the example in Figure 2.2 this number is 2.

Factor 3: Dynamic of surrounding vehicles

This factor obtains a weighted average value of the dynamic parameters of related surrounding vehicles and gives an overview of the dynamic of the scenario. The weighting factors are decided by the positions of surrounding vehicles and their movement relative to ego-vehicle. The area of ROI is divided into 12 sectors, vehicles of each sector will be assigned with a label indicating their positions within ROI (Figure 2.3). The detailed process of division can be found in [1, pp. 22-24].

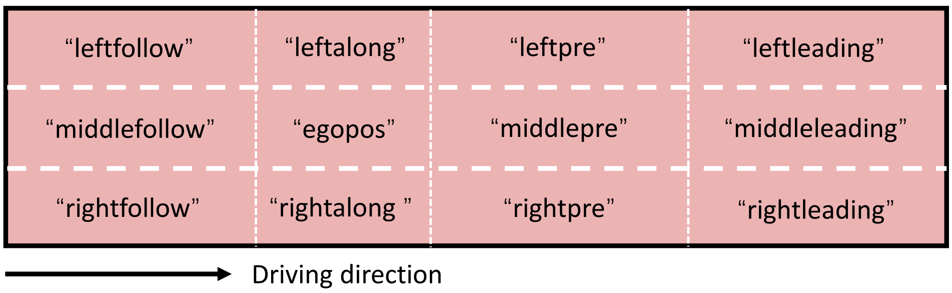


Figure 2.3: classification of surrounding vehicles

The weighting factors are determined using a case discrimination. In longitudinal direction, if the absolute value of a surrounding vehicle's velocity or acceleration is larger than ego-vehicle, weighting factors are assigned according to Figure 2.4 (a). If smaller than ego-vehicle, weighting factors from Figure 2.4 (b) will be assigned. In lateral direction, if a surrounding vehicle has a velocity or acceleration towards right, the ones on the left side of ego-vehicle will be weighted more. Vice versa, if a surrounding vehicle has velocity or acceleration towards left, the ones on the eight side of ego-vehicle are more likely to have influence and will be weighted more. Weighted average value of dynamic with respect to longitudinal velocity is then calculated with the following formula:

$$dynamic_{v,x} = \frac{\sum_{i=1}^{n} \left| v_{i,x} \right| \cdot w_{vx,i}}{\sum_{i=1}^{n} w_{vx,i}} \tag{2.2}$$

subscript vx indicates velocity in longitudinal direction, n is the total number of surrounding vehicles, w is the assigned weighting factor of the corresponding vehicle. Dynamic with respect to longitudinal acceleration, lateral velocity and acceleration are calculated similarly.

→ Driving direction

2	1	0	0
2	3/5	0	0
2	1	0	0

(a) Velocity/Acceleration of surrounding vehicles larger than ego-vehicle

0	1	2	1
0	3/5	3	2
0	1	2	1

(b) Velocity/Acceleration of surrounding vehicles smaller than ego-vehicle

Figure 2.4: Weighting factors of dynamic parameters in longitudinal direction (vehicle labelled “egopos” will be assigned with 3 if its velocity is not larger than the threshold value of 10 km/h, else with 5)

→ Driving direction

2	1	0	0
2	3/5	0	0
2	1	0	0

(a) Velocity/Acceleration of surrounding vehicles larger than ego-vehicle

0	1	2	1
0	3/5	3	2
0	1	2	1

(b) Velocity/Acceleration of surrounding vehicles smaller than ego-vehicle

Figure 2.5: Weighting factors of dynamic parameters in lateral direction (vehicle labelled “egopos” will be assigned with 3 if its velocity is not larger than the threshold value of 10 km/h, else with 5)

Factor 4: Variation of the dynamic parameters

This factor calculates the difference between the largest and smallest value of respective dynamic parameters. The variation in terms of longitudinal velocity can be expressed with the following formula:

$$variation_{v,x} = \max(\boldsymbol{v_x}) - \min(\boldsymbol{v_x}) \tag{2.3}$$

$\boldsymbol{v_x}$ is a vector containing the longitudinal velocities of all surrounding vehicles within ROI, $\max(\boldsymbol{v_x})$ is the maximum among it and $\min(\boldsymbol{v_x})$ the minimum. Variation with respect to longitudinal acceleration, lateral velocity and acceleration can be calculated with the same principle. This factor gives a hint of how different the behaviours of the surrounding vehicles are.

Factor 5: Connectivity

This factor reflects the level of mutual influence between traffic participants. The existence of vehicles in two sectors of ROI (defined in Figure 2.3) next to each other in lateral and longitudinal direction defines the existence of one connection between these two sectors. Maximal 17 connections between total 12 sectors can exist according to this definition (Figure 2.6).

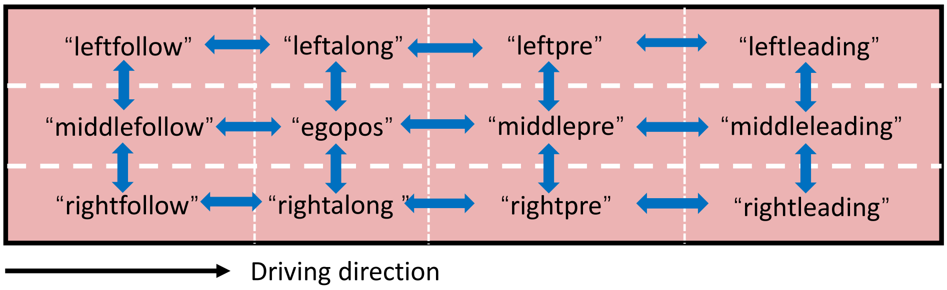


Figure 2.6: Connections within a scenario

Factor 6: Deviation of surrounding vehicles from predicted trajectories

Better prediction of surrounding vehicles' behaviour helps to improve the judgement of ego-vehicle on the scenario, which enables the automated ego-vehicle to take actions in time, so that critical situations can be avoided. Based on the results of literature research [1, pp. 31-34], the trajectories of surrounding vehicles (except vehicles labelled "leftleading", "middleleading" and "rightleading") are predicted with a constant acceleration model. The velocities of a surrounding vehicle at the actual time step (in the actual frame) in lateral and longitudinal directions are seen as the initial values of a movement with constant accelerations (accelerations at actual time step) in both directions. The time interval to be predicted is the time necessary for ego-vehicle from current velocity to come to a halt with a deceleration of 10 m/s^2 (Eq. **Fehler! Verweisquelle konnte nicht gefunden werden.**). The detailed interpretation and implementation of this factor can be found in [1, pp. 34-37].

$$t_{predict} = \frac{v_{ego,x}}{10m \cdot s^{-2}} \tag{2.4}$$

Factor 7: Possible actions of ego-vehicle

Factor 8: Possible actions of surrounding vehicles

Although Factor 6 may offer a hint of uncertainty within a scenario to a certain extent, it does not take the existence of other vehicles into consideration. Factor 7 and 8 can make up for this deficiency. These two factors will be evaluated based on the same principle. By observing the occupancy status of the eight nearest sectors around a vehicle, which can be ego-vehicle or a surrounding vehicle (in case of a surrounding vehicle, it will be temporally regarded as an "ego-vehicle" and the nearest eight sectors around it will be observed), the number of possible movement can be counted. Table 2.1 shows the possible actions under certain circumstances. As can be seen that the maximal number of possible actions is 8. A small difference between Factor 7 and 8, namely between ego-vehicle and surrounding vehicle regarding evaluation is that, the three deceleration actions will not be included for ego-vehicle. Therefore, the maximal number of possible actions for ego-vehicle is 5.

Table 2.1: Possible actions of a vehicle

Nr.	Action	Occupancy status of relevant sectors
1	Deceleration	Always possible when moving forward
2	Acceleration	No vehicle labelled “middlepre“
3	Lane change to the left	No vehicle labelled “leftalong“ and no vehicle labelled ”leftfollow” moving forward very fast (a left adjacent lane is available).
4	Acceleration after lane change to the left	Lane change to the left possible and no vehicle labelled “leftpre”
5	Deceleration after lane change to the left	Lane change to the left possible and vehicle labelled “leftpre” exists
6	Lane change to the right	No vehicle labelled “rightalong“ and no vehicle labelled ”rightfollow” moving forward very fast (a right adjacent lane is available).
7	Acceleration after lane change to the right	Lane change to the right possible and no vehicle labelled “rightpre”
8	Deceleration after lane change to the right	Lane change to the right possible and vehicle labelled “rightpre” exists

Factor 9: time-gap

This factor serves as an indicator for the question, how precisely does one action have to be executed, namely how difficult it is for ego-vehicle to perform an action. For instance, with existence of a vehicle “leftfollow”, how difficult is it for ego-vehicle to change from current lane to the left adjacent lane? A value similar to time-headway is calculated and used as a measurement for such cases. Vehicles considered for this factor are the ones labelled with “leftfollow”, “rightfollow”, “leftpre”, “middlepre”, “rightpre”. Time-gap for each of these vehicles can be expressed with the following formula:

$$time\_gap_{label} = \frac{d_{label}}{v_{ego.x}} \qquad label \in \{leftfollow, rightfollow, leftpre, middlepre, rightpre\} \tag{2.5}$$

d_{label} is the distance between ego-vehicle and a surrounding vehicle with corresponding label. For surrounding vehicles in front of ego-vehicle, this distance is the range from ego-vehicle’s front to surrounding-vehicle’s back. For surrounding vehicles behind ego-vehicle, this distance is then the range from surrounding-vehicle’s front to ego-vehicle’s back (Figure 2.7). The final value of this factor is the average value of all related surrounding vehicles. In Figure 2.7 the factor value is the average value of $time\_gap_{rightfollow}$ and $time\_gap_{leftpre}$.

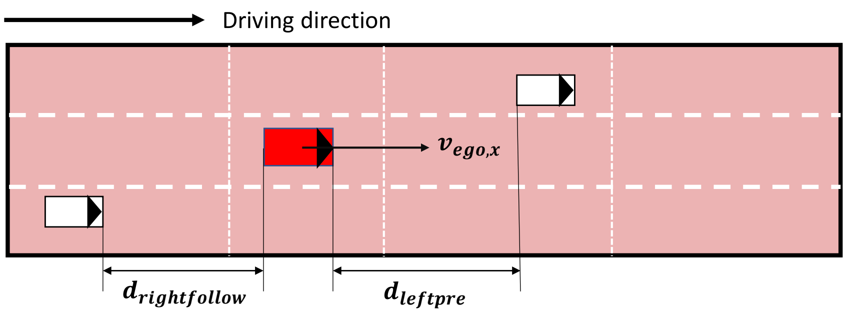


Figure 2.7: Time-gap between ego-vehicle and surrounding vehicles

Factor 10: Blind Spot

The area not reachable by various sensors of the automated ego-vehicle is defined as the blind spot. Large area of blind spot rises the uncertainty of the scenario and makes it difficult for ego-vehicle to deal with unexpected situations. Various sensors, which are supposed to be installed at different positions of a vehicle, are assumed to be installed at only two positions for simplification. One is in the middle of ego-vehicle's front and is for the detection of surrounding vehicles in front of ego-vehicle's midpoint. The other is in the middle of ego-vehicle's back and is for the detection of surrounding vehicles behind ego-vehicle's midpoint (Figure 2.8). The algorithm for the calculation of these discrete blind spot areas can be found in [1, pp. 42-48].

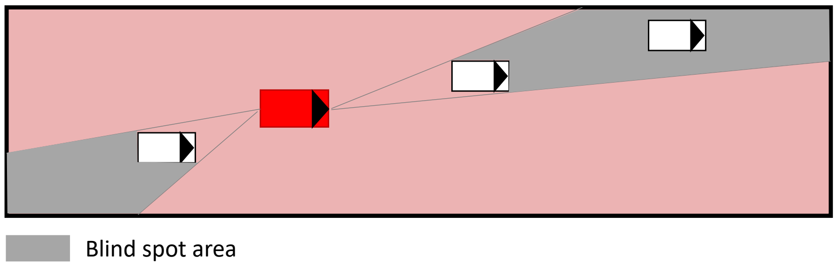


Figure 2.8: Blind spot area within ROI

2.1.2 Normalization and Complexity

In the last section it is introduced, how each factor is defined and calculated. The values of these factors can lie in very different ranges. For instance, types of surrounding vehicles can be from 0 to 2, number of surrounding vehicles can be from 0 to larger than 10, the variation of dynamic parameters of surrounding vehicle can range from 0 to about 15 m/s, etc. In addition to different ranges, the units of these values are different as well. Therefore, it is necessary to normalize these values, so that they can be brought to the same scale for the calculation of complexity. How the value of each factor is normalized is listed in Table 2.2. Variables with subscript "value" indicate the actual value of influence factors. Subscript "norm" (or variables on the left side of the equation) indicate the normalized values and "ref" the reference value used for normalization.

Table 2.2: Normalization of influence factors

Nr.	Influence factor	Normalization (Scene-based value)
1	Types of surrounding vehicles	$type_{norm} = \frac{type_{value}}{type_{ref}}, \quad type_{ref} = 2$
2	Number of surrounding vehicles	$veh_{norm} = \frac{veh_{value}}{veh_{ref}}, \quad veh_{ref} = 11$
3	Dynamic of surrounding vehicles	$dynamic_{norm} = \frac{1}{4}(dynamic_{norm,vx} + dynamic_{norm,vy} + dynamic_{norm,ax} + dynamic_{norm,ay})$, with: $dynamic_{norm,vx} = \frac{dynamic_{value,vx}}{dynamic_{ref,vx}},\ dynamic_{ref,vx} = 35\ m/s,$ $dynamic_{norm,vy} = \frac{dynamic_{value,vy}}{dynamic_{ref,vy}},\ dynamic_{ref,vy} = 0.65 m/s,$ $dynamic_{norm,ax} = \frac{dynamic_{value,ax}}{dynamic_{ref,ax}},\ dynamic_{ref,ax} = 0.65\ m/s^2,$ $dynamic_{norm,ay} = \frac{dynamic_{value,ay}}{dynamic_{ref,ay}},\ dynamic_{ref,ay} = 0.22\ m/s^2.$
4	Variation of dynamical parameters	$variation_{norm} = \frac{1}{4}(variation_{norm,vx} + variation_{norm,vy} + variation_{norm,ax} + variation_{norm,ay})$, with: $variation_{norm,vx} = \frac{variation_{value,vx}}{variation_{ref,vx}},\ variation_{ref,vx} = 15\ m/s,$ $variation_{norm,vy} = \frac{variation_{value,vy}}{variation_{ref,vy}},\ variation_{ref,vy} = 1.3\ m/s,$ $variation_{norm,ax} = \frac{variation_{value,ax}}{variation_{ref,ax}},\ variation_{ref,ax} = 1.5\ m/s^2,$ $variation_{norm,ay} = \frac{variation_{value,ay}}{variation_{ref,ay}},\ variation_{ref,ay} = 0.5\ m/s^2.$
5	Connectivity	$Connectivity = \frac{connection_{value}}{connection_{ref}}, \quad connection_{ref} = 17$
6	Deviation from predicted trajectory	$deviation_{norm} = \frac{deviation_{value}}{deviation_{ref}}, \quad deviation_{ref} = 1.4\ m$
7	Possible actions of ego-vehicle	$pa_{norm,ego} = \begin{cases} 0 & pa_{value,ego} = 0 \\ \frac{1}{4}(-pa_{value,ego} + 5) & pa_{value,ego} > 0 \end{cases}$
8	Possible actions of surrounding vehicles	$pa_{norm,sur} = \begin{cases} 0 & pa_{value,sur} = 0 \\ \frac{1}{7}(-pa_{value,sur} + 8) & pa_{value,sur} > 0 \end{cases}$
9	Time gap	$time\_gap_{norm} = e^{-0.4 \cdot time\_gap_{value}}$
10	Blind spot	$ratio = \frac{S_{blind\_spot}}{S_{ROI}}$

The complexity of each scene is achieved by linear combination of the normalized values of these influence factors. Each influence factor will be equally weighted and their weighting factors add up to 1.

$$\mathrm{C_{scene}} = \sum_{i=1}^{10} w_i \cdot f_i\,, \qquad w_i = \frac{1}{10} \tag{2.6}$$

w_i and f_i represent the weighting factor and normalized value of factor i. There are 10 factors in total. $\mathrm{C_{scene}}$ is the complexity of a scene. Vector $\mathbf{C_{scene}}$ contains the complexity of every scene belonging to the same scenario. Complexity of the scenario $\mathrm{C_{scenario}}$ is defined as the average complexity of all scenes in this scenario:

$$C_{scenario} = mean\ (\boldsymbol{C_{scene}}) \tag{2.7}$$

2.2 Sensitivity analysis

In the previous work the final results of complexity are obtained by linear combination of normalized values of influence factors with equal weight. The problem of equal weighting factors has already been mentioned in section 1.1. On the one hand, it is unrealistic that all factors are equally important. For instance, factors which reflex the uncertainty or criticality of the scenario should be weighted more compare with other factors. On the other hand, since the normalized values of different factors have different ranges, for some 0.5 is quiet large value, while for others 0.5 is only in a medium level. Thus, although weighting factors are the same, the factors are actually not equally weighted due to different ranges of their values. Based on these two considerations, it is necessary to make adjustments to the weighting factor of each influence factor, so that the scenarios selected through this metric can be more convincing and more representative.

Deciding the relative importance of each influence factor can be a subjective process, which will be discussed in detail in the next section. The results are on the one hand influenced by the weighting factors, and on the other hand influenced by the values of influence factors. Since different values have different distributions, changes in different factors can cause different influences on the result of complexity. This kind of influence can be studied objectively with the so called "Sensitivity Analysis", whose definition according to the work of SALTELLI ET AL. [5, p. 1]is the study of how uncertainty in the output of a model is distributed among or caused by the uncertainty of different inputs of this model (the model can be numerical or otherwise). Before conduction of sensitivity analysis, an uncertainty analysis of the model needs to be carried out first. One possibility for such analysis is the Monte Carlo analysis [5, pp. 6-7], which will be explained in detail in the following text.

$x_1, x_2, \ldots., x_m$ are the inputs of the system, m is the total number of inputs. Their distributions are expressed in $N(\overline{x_1}, \sigma_{x_1})$, $N(\overline{x_2}, \sigma_{x_2})$,…, $N(\overline{x_m}, \sigma_{x_m})$, with $\overline{x_1}$, $\overline{x_2}$,…, $\overline{x_m}$ the mean values and σ_{x_1}, σ_{x_2},…, σ_{x_m} the standard deviations of respective inputs. For simplification it is assumed that these inputs are independent of each other. A sample or a set of inputs can be obtained by drawing one element from each distribution. n sets of inputs can be expressed with a n x m matrix (Figure 2.9). With each set of inputs $(x_1^{(i)}, x_2^{(i)}, \ldots, x_m^{(i)})$ an output $y^{(i)}$ $(i = 1,2, \ldots, n)$ can be calculated. Furthermore the average value, standard deviation, distribution, etc of outputs of the model can be obtained as well.

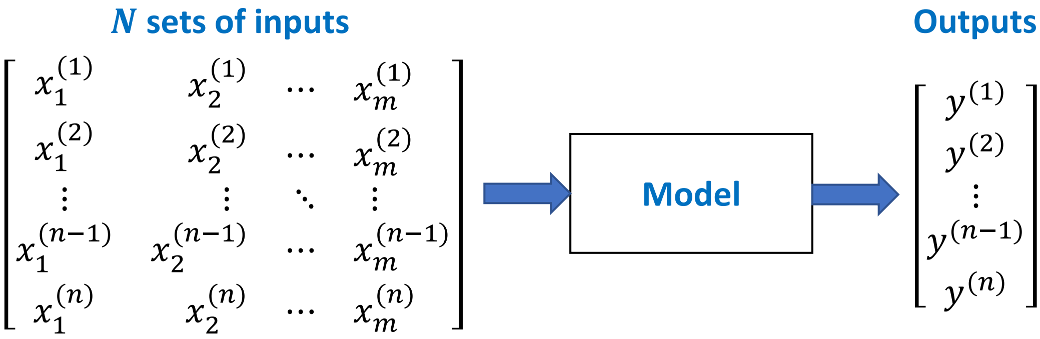


Figure 2.9: Uncertainty analysis (Monte Carlo analysis)

Based on the results of uncertainty analysis, the sensitivity analysis can then be performed in several ways. Since the model used in this thesis is linear, these methods will be introduced with help of a linear model (Eq.(2.8)).The methods introduced in the rest part of this section can all be refered to SALTELLI ET AL. [5, pp. 10-20]'s work.

$$y = \sum_{i=1}^{m} w_i \cdot x_i \tag{2.8}$$

y is a single output, $x_1, x_2, \dots, x_m$ are the m input factors, $w_1, w_2, \dots, w_m$ are coefficients which can be for instance system characteristic parameters or weighting factors. At the same time these methods are based on the assumption that the input factors are independent from each other.

Scatterplots

Without need for further computation after uncertainty analysis, a scatterplot can be created for each input factor. Taking input x_1 for example, its scatterplot can be created by plotting points with coordinates $\left(x_1^{(1)}, y^{(1)}\right), \left(x_1^{(2)}, y^{(2)}\right), \dots, (x_1^{(n)}, y^{(n)})$ in a coordinate system. The influence level of each input can be determined by observing the shape of respective point cloud. If the shape of point cloud is close to a circle, then this input has little influence on output. If the point cloud shows a linear relationship between this input factor and the output, the stronger this relationship is, the more influential is this input factor.

Derivative

Derivative is a common and simple way used to describe how sensitive is the change of output caused by a change in input. The result can be seen as a measure for relative importance of respective input factor x_i (Eq.(2.9)).

$$S_{x_i}^{p} = \frac{\partial y}{\partial x_i}, i = 1,2, \dots, m \tag{2.9}$$

p stands for "partial derivative". For the model mentioned in Eq. (2.8), the input factors are supposed to be equally important according to the result of this method, since their weighting factors are all equal to w. However, according to the assumption these input factors have different values of standard deviation. Therefore, it is not convincing to make the conclusion that they are equally important.

Sigma-normalized Derivative

To make up for the shortcoming of derivative, sigma-normalized derivative is developed by taking standard deviation of respective input factor into consideration (Eq.(2.10)). Thus, the original derivative $S_{x_i}^{p}$ is weighted and normalized by the ratio of standard deviation between input and output.

$$S_{x_i}^{\sigma} = \frac{\sigma_{x_i}\partial y}{\sigma_y \partial x_i}, i = 1,2, \ldots, m \tag{2.10}$$

If all input factors are standardized (the input is subtracted by its mean value and then divided by its standard deviation), there exists the following relationship between standard deviation of output and standard deviation respective input factor:

$$\sigma_y^2 = \sum_{i=1}^{m} w_i^2 \sigma_{x_i}^2 \tag{2.11}$$

By replacing $\frac{\partial y}{\partial x_i}$ in Eq. (2.10) with w_i and replacing σ_y in Eq. (2.11) with $\frac{\sigma_{x_i}}{S_{x_i}^{\sigma}} w$, following conclusion can be obtained. The value of $(S_{x_i}^{\sigma})^2$ offers information about how much the input factor x_i contributes to the variance of the output. If every input factor has a normal distribution with 0 as the average value. $(S_{x_i}^{\sigma})^2$ of all input factors will add up to 1:

$$\sum_{i=1}^{m} (S_{x_i}^{\sigma})^2 = 1 \tag{2.12}$$

Standardized Regression Coefficient

Another possibility is the application of linear regression. By calculating the squared difference between the output of linear regression model and the output obtained by uncertainty analysis the coefficients k_0 and k_i can be determined (least-square).

$$y(j) = k_0 + \sum_{i=1}^{m} k_i\, x_i^{(j)}, j = 1,2, \ldots, n \tag{2.13}$$

Standardized regression coefficient (SRC) β_{x_i} is defined as $k_i \frac{\sigma_{x_i}}{\sigma_y}$. Since the actual outputs are obtained with linear model, therefore, if n is large enough the value of β_{x_i} should be almost the same as that of $S_{x_i}^{\sigma}$. Similar to sigma-normalized derivative, if all input factors has a normal distribution with an average value 0, the sum of squares of SRC is equal to 1 as well:

$$\sum_{i=1}^{m} (\beta_{x_i})^2 = 1 \tag{2.14}$$

For linear models, SRC has the same effect compared to sigma-normalized coefficients. The difference is that, SRC shows better robustness and reliability when applied to nonlinear models, since compared to sigma-normalized derivative SRC is multidimensional (However, if n is not large enough and the number of inputs is large, SRC would not be a very precise method).

3 Method

In this chapter, several deficiencies in the previous work will be improved. The metric will be extended with three more factors, which are not included in the previous work, but might have influence on the complexity as well. The main goal of this chapter is to study the relationship between influence factors and complexity. A sensitivity analysis will be conducted with equal weighting factors to evaluate the degree of influence of each factor on complexity. A questionnaire will be desgined so that the knowledge and experience of experts can be taken into consideration when determining the weighitng factors of each factor. An overview of the content and process of the work is show in the following figure.

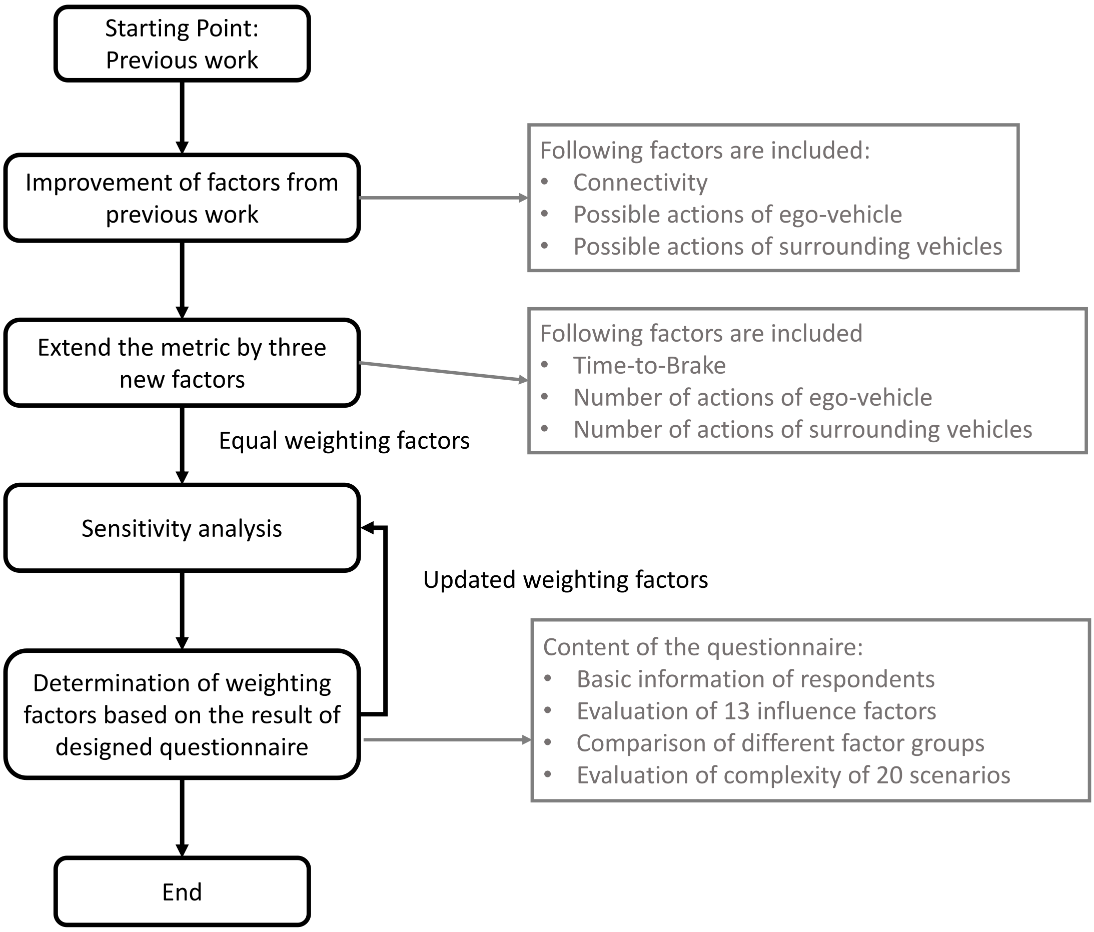


Figure 3.1: Flow chart of the method

3.1 Improvement of the previous work

In this section some shortcomings in the previous work will be corrected, which involves factor "Connectivity", "Possible actions of ego-vehicle" and "Possible actions of surrounding vehicles".

3.1.1 Connectivity

One major change to be made is the definition of connectivity. Previously, connectivity is defined as the total number of connections, for which only the ones in lateral and longitudinal direction are taken into consideration. This consideration is incomplete for the real-world traffic. For the case that ego-vehicle intends to change to the left or right adjacent lane, or that vehicles labelled "leftpre" or "rightpre" can possibly cut in in front ego-vehicle, the connections between ego-vehicle and surrounding vehicles in diagonal directions need to be considered as well (Figure 3.2).

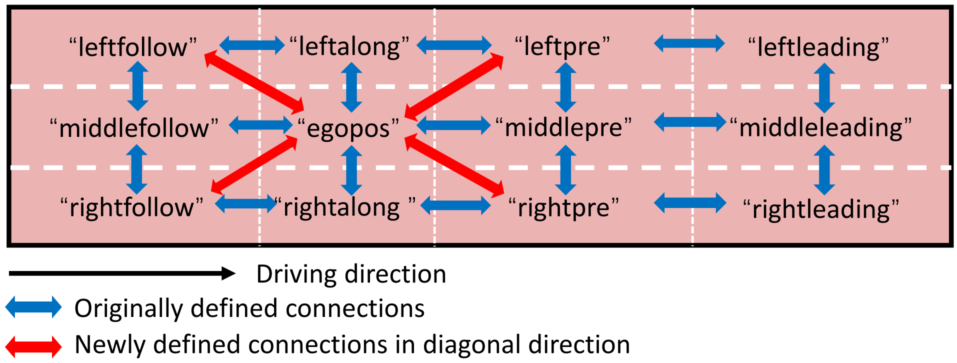


Figure 3.2: Connections after improvement

In the previous definition, a connection will be counted, as long as the sectors next to each other in longitudinal direction or lateral direction are occupied, regardless of the number of vehicles in each sector. However, in the reality, it is possible that there is more than one vehicle in one sector, which can all have influence on vehicles from the neighboring sector. For more accuracy, the total number of connections will consist of the following four parts: First, if there are s_i vehicle(s) within sector i $(i = 1,2, \dots, 12)$, the number of connections within sector i will be $s_i - 1$. For all sectors, this number will be $\sum_{i=1}^{12}(s_i - 1)$. Second, for connections between vehicles from two adjacent sectors in lateral direction, situations will be divided into four categories shown in Figure 3.3. If one of the two sectors are vacant, number of connections in lateral direction in this case will be 0 (Figure 3.3(1)). If each one of the two sectors have exactly one vehicle, number of connections will be 1 (Figure 3.3(2)). If one sector is occupied with one vehicle and the other with two, the number of connections will be 2 (Figure 3.3(3)). One sector occupied with more than two vehicles is very unlikely in the reality. Finally, if both sectors are occupied with two vehicles, the number of connections will be 4 (Figure 3.3(4)). In general, number of connections between two adjacent sectors can be expressed as $s_i \cdot s_j$ (s_i and s_j are the numbers of vehicles of two laterally adjacent sectors i and j). Third, the calculation of connections between two longitudinally adjacent sectors remain unchanged, namely if both sectors are occupied, regardless of the number of vehicles, one connection will be counted. The reason for this is that, for cases with more than one vehicle in a sector, the connections within sector has already been counted, the connection between sectors need to be counted only once, since one vehicle does not have direct interaction with a vehicle behind/in front of its following/preceding vehicle (Figure 3.4). The fourth part are the connections in diagonal direction. If there is a vehicle with label “leftfollow”, “rightfollow”, “leftpre” or “rightpre” (namely within the sector in diagonal direction of ego-vehicle), a connection will then be counted. The number of vehicles within one sector does not play a role here, for case with two vehicles within one sector, only the one closer to ego-vehicle can have large enough influence. Therefore, there are maximal 4 connections in diagonal direction.

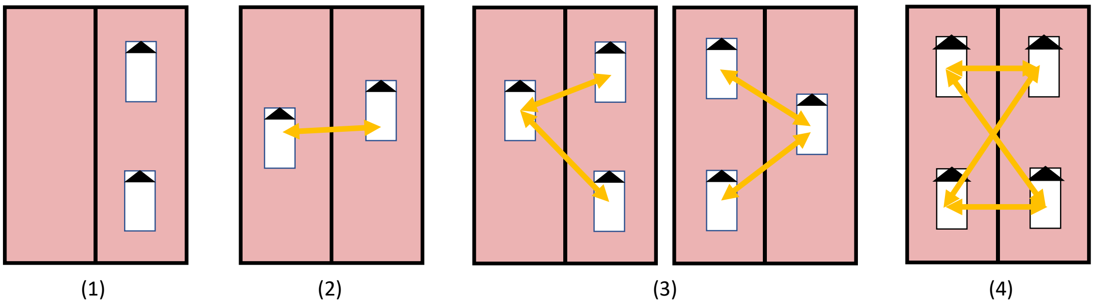


Figure 3.3: Definition of connections in lateral direction

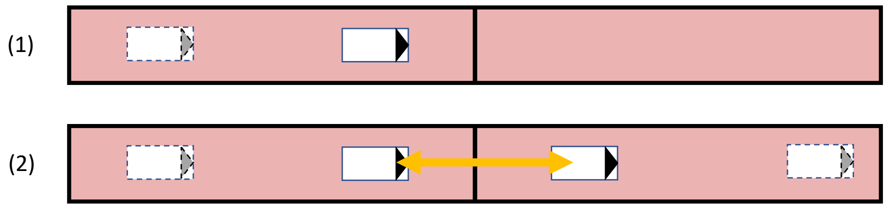


Figure 3.4: Definition of connection in longitudinal direction

When implementing the second and third part of the connections (connections between sectors) with MATLAB, each two adjacent sectors will be seen as a group (Figure 3.3 and Figure 3.4). In lateral direction there are 8 (2 x 4) such groups in total. The number of connections of each group will be examined and expressed as $c_{lat.m}$ $(m = 1,2,..,8)$. The total number of connections in lateral direction adds up to $\sum_{m=1}^{8} c_{lat,m}$. In longitudinal direction there are 9 (3 x 3) such groups in total. The number of connections of each group is expressed as $c_{lon.n}$ $(n = 1,2,..,9)$. The total number of connections in the longitudinal direction adds up to $\sum_{n=1}^{9} c_{lon,n}$. Together with the connections in diagonal directions c_{diag} and the connections within sectors $\sum_{i=1}^{12}(s_i - 1)$, the total number of connections within the entire ROI can be expressed with the following formula:

$$connection_{value} = \sum_{i=1}^{12}(s_i - 1) + \sum_{m=1}^{8} c_{lat,m} + \sum_{n=1}^{9} c_{lon,n} + c_{diag} \tag{3.1}$$

Since the possibility of two vehicles appearing in one sector is relatively small, the reference value used for normalization will be increased by 4 from the original value, which corresponds to the four newly added connections between the ego-vehicle and surrounding vehicles in diagonal direction. $connection_{ref}$ is now equal to 21.

3.1.2 Possible actions of ego-vehicle and surrounding vehicles

The other factors being improved are possible actions of ego-vehicle and of surrounding vehicles. The maximal number of possible actions of ego-vehicle is 5 and of a surrounding vehicle is 8. In the previous work, these two factors are normalized in such a way, that the normalized values are zero for no possible actions and 1 for there is only one possible action. The argumentation for this is that, situations are supposed to be simple for two extreme cases, namely there is no surrounding vehicles and the vehicle (in this case the vehicle can be ego-vehicle or a surrounding vehicle) has all the possibilities, or all sectors around the vehicle are occupied with other vehicles and it has no option for other movement but to maintain the current state. For these two cases the factors values should be zero after normalization. The largest normalized value is supposed to appear at a position where a vehicle has a particular number of possibilities.

According to the interpretation of these factors in the previous work [1, p. 38], these factors should “reflect the indistinction of the target situation”, which is one of the characteristic of complex scenario. However, from the perspective of ego-vehicle, this characteristic is more reflected by the possible actions of surrounding vehicle. The behaviors of surrounding vehicles will be difficult to predict if they have large number of possible actions, which brings more ambiguity for ego-vehicle. The possible actions of ego-vehicle itself more reflects the effort with respect to decision making for the current traffic situation. If ego-vehicle has no possible actions or all the possibilities, the decision making would be an easier process. Based on these considerations, the normalization of the values of these two factors will be improved. A monotone increasing function will be used for the normalization of possible actions of surrounding vehicles, so that larger number of possible actions of surrounding vehicle reflects more ambiguity of the scenario for ego-vehicle (Eq. (3.2)).

$$pa_{norm,sur} = \frac{1}{8} pa_{value,sur}, \qquad pa_{value,sur} = 1,2,\dots,8 \tag{3.2}$$

$pa_{value,sur}$ represents the average value of number of possible actions of surrounding vehicles and $pa_{norm,sur}$ is the normalized value of $pa_{value,sur}$.

For normalization of possible actions of ego-vehicle a piecewise function is chosen, so that the two extreme cases mentioned above indicate small effort for decision making for ego-vehicle and medium range of possible actions requires more effort.

$$pa_{norm,ego} = \begin{cases} \frac{1}{2} pa_{value,ego}, & pa_{value,ego} = 0,1,2 \\ -\frac{1}{2} pa_{value,ego} + \frac{5}{2}, & pa_{value,ego} = 3,4,5 \end{cases} \tag{3.3}$$

$pa_{value,ego}$ is the number of possible actions of ego-vehicle and $pa_{norm,ego}$ the normalized value of $pa_{value,ego}$. With this method, the maximal normalized value 1 appears when ego-vehicle has 2 or 3 possible actions (Figure 3.5).

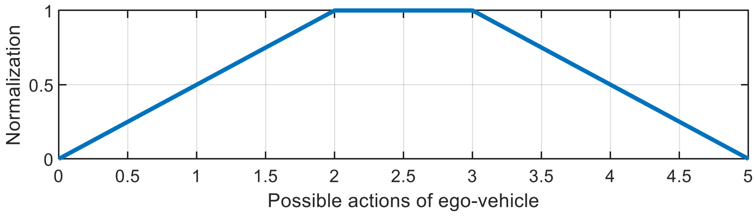


Figure 3.5: Normalization of possible actions of ego-vehicle

3.1.3 Time-gap

In the previous work, time-gap is defined as the longitudinal distance between ego-vehicle and the surrounding vehicle (with label “leftfollow”, “rightfollow”, “leftpre”, “middlepre” or “rightpre”) divided by the longitudinal velocity of ego-vehicle. If the corresponding position of a surrounding vehicle is vacant, then the related time-gap has a default value of 6 s. The value $time\_gap_{value}$ used for normalization in Table 2.2 is actually the average value of all five vehicles with labels mentioned above. In most cases, only one or two of the five sectors are occupied, for other empty sectors time-gap has the default value 6. This will lead to a large $time\_gap_{value}$ after averaging and a very small normalized value. Therefore, assuming $\boldsymbol{time\_gap_{value}}$ is a vector containing time-gap value of all five sectors, only the maximal value will be used for the normalization. The expression in Table 2.2 can be expressed as:

$$time\_gap_{norm} = e^{-0.4 \cdot \max(\boldsymbol{time\_gap_{value}})} \tag{3.4}$$

3.2 Extension of the previous work

In this section three factors which are not included in the previous work will be introduced and taken into consideration for the evaluation of complexity.

3.2.1 Time-to-Brake

In the previous work it has been distinguished between the two concept “complexity” and “criticality”. If a scenario is critical for ego-vehicle, which means that ego-vehicle is on the verge of an accident. The ten influence factors are derived based on the characteristics of “complexity”. A complex scenario can be critical at the same time or it can evolve into a critical situation in a time. Namely, these two concepts are not mutually exclusive. However, criticality has not been included in the previous work. Therefore, an influence factor reflecting the scenario’s criticality or the potential of getting critical will be complemented to the previous work.

In NOH ET AL.’s work [6], three values are used as “threat measures” to evaluate the possibility of collision. The first value is time-to-collision, which is calculated from the distance between two vehicles, divided by the speed difference of the two vehicles. The second value is the safety distance ought to be kept, so that minimum safety can be ensured. The third value being used is time-to-brake, which is defined as the time remaining for ego-vehicle to take emergency brake to avoid collision in case the preceding vehicle suddenly comes to a standstill.

In JUNIETZ’s work [7], a Trajectory Criticality Index (TCI) was introduced for the assessment of criticality, which consists of three elements: acceleration, reaction time and precision. Acceleration is selected as an element because this maneuver is usually inevitable in a critical situation. Normalization of acceleration is achieved with help of Kamm’s circle. Under the premise that the coefficient is known. Acceleration will be normalized by dividing it by the maximum achievable acceleration $\mu \cdot g$, with μ the friction coefficient and g the gravitational constant. Since the friction coefficient can only be estimated for many reasons, instead of a friction circle, other shapes like a rhombus or a cross can be used to ensure that the maximum achievable acceleration will not be exceeded (the boundaries of other shapes lie within the friction circle). The second element reaction time takes two situations into consideration: in the first situation where no evasion maneuver is possible, only time-to-brake in longitudinal direction will be considered. In the second situation where there is possibility to evade the collision, time-to-steering in lateral direction will be calculated. The final reaction time is the larger value of time-to-brake and time-to-steering. The normalized value is achieved by using a reference value of 2 s and a monotone decreasing function. The third element precision observes the dynamic in lateral direction. The reaction time for this component consists of two parts: the first part is the time necessary for ego-vehicle to catch up with the preceding vehicle, the second part is obtained by dividing lateral distance between two vehicles with the lateral velocity of ego-vehicle. The value for this element is normalized with help of a reference value of 2 s and a monotone decreasing function as well.

The three elements used for TCI are based on the human driver experience. Some measures used by the research mentioned above are not very appropriate for automated vehicles. An automated vehicle should be able to keep a safety distance to the preceding vehicle automatically with help of, for instance, an ACC system. Therefore, it is not likely to get into a critical situation due to not keeping the safety distance. The first two measures of NOH ET AL.'s work [6] will not be considered. In other words, the criticality of a scenario from the perspective of an automated ego-vehicle should mainly be caused by the unexpected or unpredicted behaviors of surrounding vehicles, not the improper behaviors of ego-vehicle. Since friction coefficient is not available in HighD dataset, element "Acceleration" from JUNIETZ's work [7] will not be considered as well. The element "Precision" in TCI has already been taken into consideration as a separate influence factor "time-gap" in the previous work [1, pp. 40-41]. Based on the consideration above, time-to-brake will be selected in this thesis as a newly added influence factor, which indicates the criticality of a scenario.

For this factor not all vehicles in front of ego-vehicle will be observed, only the one labelled with "middlepre" (if there are two vehicles have this label, only the one closer to ego-vehicle will be considered). If ego-vehicle has a larger velocity in longitudinal direction than the preceding vehicle, the time will be calculated, which ego-vehicle needs to decelerate until it has the same velocity as the preceding vehicle with constant maximum deceleration a_{max} (-10 m/s^2). $d_{middlepre}$ represents the distance between ego-vehicle's front and the preceding vehicle's rear. The distance covered during movement with constant deceleration can be expressed with the following formula:

$$d_{dec} = \frac{v_{ego,x}^2 - v_{middlepre,x}^2}{2 \cdot |a_{max}|} \tag{3.5}$$

$v_{ego,x}$ and $v_{middlepre,x}$ represent the longitudinal velocity of ego-vehicle and vehicle "middlepre" respectively. For instance, if ego-vehicle moves with a constant velocity of 30 m/s and vehicle "middlepre" stands in still. Then ego-vehicle needs 45 m to come to a standstill. It is noticeable, that this distance is only a bit shorter than the safety distance used to determine the area of ROI, which is 54 m (30 m/s x 1.8 s). The time which remains for ego-vehicle to take the decision is then:

$$ttb_{value} = \frac{d_{middlepre} - d_{dec}}{v_{ego,x}} \tag{3.6}$$

The result can be negative, in which case a critical situation is considered to be inevitable even ego-vehicle takes an emergency brake. The corresponding normalized value in this case will be 1. A reference value of 2 s is used for normalization when ttb_{value} is positive. If $ttb_{value} > 2s$, the corresponding normalized value will be 0. If ttb_{value} is between 0 and 2, like in JUNIETZ's work [7], a monotone decreasing function will be used for normalization (Eq. (3.7)).

$$ttb_{norm} = \begin{cases} 1 & ttb_{value} < 0 \\ 0 & ttb_{value} > 2 \\ 1 - \dfrac{ttb_{value}}{2} & 0 \leq ttb_{value} \leq 2 \end{cases} \tag{3.7}$$

ttb_{norm} represents the value after normalization. This value will be calculated for each scene of the scenario.

3.2.2 Number of actions of ego-vehicle and surrounding vehicles

The factors derived in the previous work and the newly added factor in section 3.2.1 are only scene-based, which means that the factors are calculated for each scene of a scenario. The value of each factor is usually the average value, maximum or minimum of all scenes, therefore, reflects the average level or extreme level of different aspects of a scenario. There has not been a factor which is scenario-based and reflects the characteristic of the scenario. Based on this consideration, the two factors "number of actions of ego-vehicle and surrounding vehicles" are included for the evaluation of scenario complexity.

In SCHÖRNER's work [8] several action templates are defined to describe the development of state of a vehicle. The different action templates distinguish from each other according to different ranges of acceleration they are in, from strong deceleration to strong acceleration. This method of description will be adapted in this section for the evaluation of new factors. For ego-vehicle and surrounding vehicles, it will be distinguished between number of actions in longitudinal direction and in lateral direction. In lateral direction the number of actions is the number of lane changes, it can be obtained by counting how many times the lane ID of a vehicle has changed during the scenario. For longitudinal direction it is necessary to first determine a time interval, in which an action template or a driving state Is defined. The number of actions in longitudinal is then the number of times a vehicle's driving state changes.

HighD dataset has a frame rate of 25 fps. The time interval between two consecutive frames is 0.04 s. In other words, the situation in the road section is captured every 0.04 s. For the definition of driving state, a time interval of 0.4 s can be obtained when taking 10 consecutive frames as a unit. The acceleration of the driving state can be calculated from the average value of longitudinal accelerations of these 10 frames. n_f is the total number of frames of a vehicle (ego-vehicle or surrounding vehicle). The number of states can be calculated by dividing n_f by 10 and applying ceiling function to the result of division (Eq. (3.8)).

$$n_{state} = \lceil n_f/10 \rceil \tag{3.8}$$

n_{state} represents the number of states. Ceiling function maps the input to an integer, which is the smallest integer larger than input. For instance, a vehicle has in total 323 frames, this value divided by 10 results in 32.3. n_{state} is supposed to be the smallest integer larger than 32.3, n_{state} is equal to 33. A visualization of this process is shown in the following figure, that the movement of a vehicle in a scenario is divided into a series of driving states.

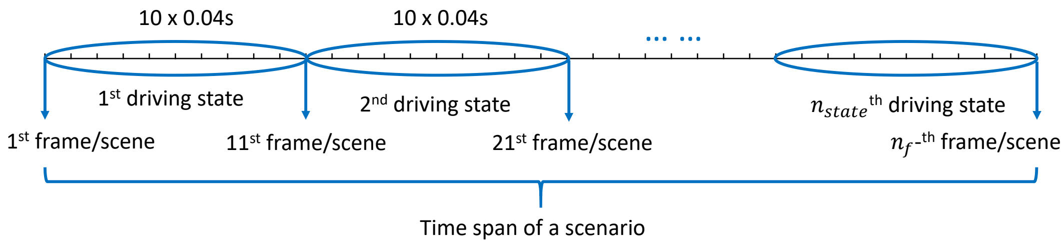


Figure 3.6: Definition of a driving state

Since in HighD dataset the longitudinal acceleration of each vehicle in each frame is available, the longitudinal acceleration of a state s_i $(i = 1,2, \dots, n_{state})$ can be obtained by calculating the mean value of accelerations of all frames contained in state s_i (Eq. (3.9)).

$$a_{s_i} = \begin{cases} \dfrac{a_{s_{i,1}} + a_{s_{i,2}} + \cdots + a_{s_{i,10}}}{10}, & i = 1,2, \dots, n_{state} - 1 \\ \dfrac{a_{s_{i,1}} + a_{s_{i,2}} + \cdots + a_{s_{i,j}}}{j}, & i = n_{state}, j \leq 10 \end{cases} \tag{3.9}$$

$s_{i,j}$ is the j-th frame of state s_i. Since $n_f/10$ is not always an integer, in other words, the number of frames contained in the last state $s_{n_{state}}$ can sometimes be less than 10. For the example mentioned above, the last state of the vehicle with 323 frames has only 3 frames. In this case, the acceleration of last state will be the mean value of these 3 frames.

Several thresholds are defined, so that if a_{s_i} lies in a certain range, the state will be characterized correspondently. If a_{s_i} is larger than -0.2 m/s^2 and smaller than 0.2 m/s^2, the vehicle is considered to be in a state where it moves with a nearly constant velocity. If a_{s_i} is larger than 0.2 m/s^2 and smaller than 2 m/s^2, than it is in a state where it moves with a normal acceleration. If a_{s_i} is larger than 2 m/s^2, then the vehicle is considered to be experiencing a strong acceleration. If a_{s_i} is smaller than -0.2 m/s^2 and larger than -3 m/s^2, the vehicle is in a state with normal deceleration. If a_{s_i} is smaller than -3 m/s^2 and larger than -6m/s^2, the level of deceleration is categorized as strong. If a_{s_i} is smaller than -6 m/s^2, the vehicle is probably running into a critical situation and taking an emergency brake. A summary can be seen in the table below:

Table 3.1: Definition of a vehicle'

Range of a_{s_i}	**Driving state s_i (state code)**
$a_{s_i} \leq -6m/s^2$	Emergency brake (-3)
$-6m/s^2 < a_{s_i} \leq -3m/s^2$	Strong deceleration (-2)
$-3m/s^2 < a_{s_i} < -0.2m/s^2$	Normal deceleration (-1)
$-0.2m/s^2 \leq a_{s_i} \leq 0.2m/s^2$	Constant driving (0)
$0.2m/s^2 < a_{s_i} < 2m/s^2$	Normal acceleration (1)
$2m/s^2 < a_{s_i}$	Strong acceleration (2)

It is noticeable that categorization of driving states for acceleration and deceleration is not symmetric, the absolute value of deceleration considered as strong level is much larger than that of acceleration. The reason for this difference is that, larger deceleration can more easily be achieved by lightly pressing the brake pedal when compared with acceleration.

For better visualization of the development of driving state of a vehicle, each type of state is assigned an integer as its code (in the brackets behind each type of state). An example is shown in Figure 3.7, which is from vehicle 110 in the 1st track. The blue line in the graph is the development of longitudinal acceleration. Each red dot represents a type of driving state of 10 consecutive frames. As can be seen that, the first 4 red dots have the value 1, which corresponds with the state “normal acceleration”. The following red dots have the value 0, which means that the vehicles moves with nearly constant velocity, In the last roughly 40 frames, the acceleration is smaller than -0.2m/s^2, this state is reflected by the last 4 red dots with value -1. The vehicle has experienced change of state twice. Namely from “normal acceleration” to “constant driving” and from “constant driving” to “normal deceleration”. An action is defined as one change of state, then in this scenario the number of actions of this vehicle is 2.

Figure 3.8 shows the distribution of number of actions of all vehicles from the 1st track. The vehicles from the 1st track have on average about 300 frames, which is a scenario about 12 seconds. As can be seen from the graph that, most vehicles have less than 5 actions, this can be used as a reference value for normalization. However, the situation can be quite different when looking at the results in the 25th track, in which the number of actions of a vehicle can sometimes be larger than 20. In this case, 5 will be no longer suitable as a reference value for normalization. For the vehicles in the 1st track, assuming it takes them on average 12 seconds to pass a road section of 420 m (length of road section recorded by drone in HighD dataset), the average speed of these vehicles can be up to 35 m/s. While most vehicles in the 25th track have a velocity only half of this value or even smaller, it takes them twice as much or more time to pass the same road section, which results in larger number of frames, larger number of driving states and larger number of actions. This is also a reasonable result, since if the vehicles move with a small velocity on the highway, it is very likely that the vehicles are in a traffic jam. Situation like this often involves the stop-and-go type of movement, frequent braking, starting, acceleration and deceleration will then result in frequent changes of driving states, which then leads to large number of actions.

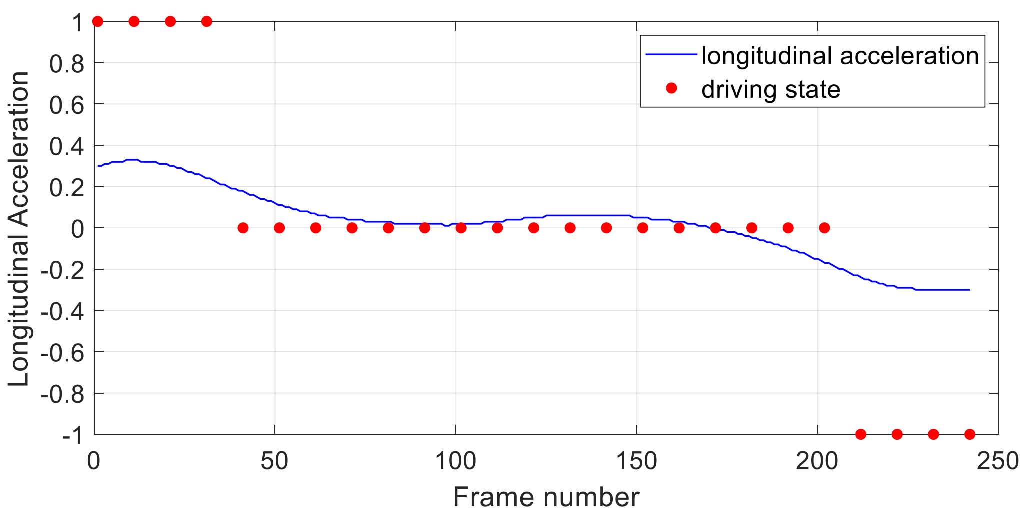


Figure 3.7: Development of driving state of vehicle Nr.110 from track 1

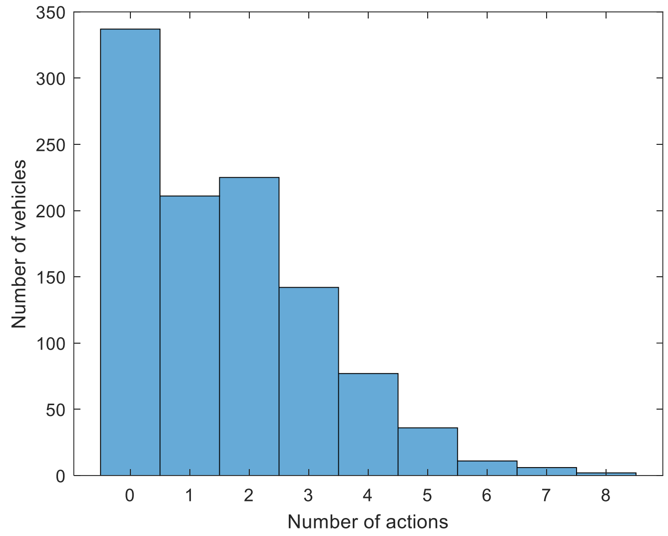


Figure 3.8: Distribution of number of actions of vehicles from the 1st track

Since large number of actions is normal for traffic flow with small velocity, so it does not necessarily relate to very complex situations. Therefore, for a reasonable normalization of this value number of frames should be taken into consideration. The solution to this problem is to define the reference value proportionally according to the number of frames. Table 3.2 shows the values of this factor of several tracks (only vehicles in negative driving direction). Column "Nr" contains the track number. "$\overline{nof}$" contains the average number of frames of all vehicles from corresponding track. "ego.x" represents the number of actions of ego vehicle in longitudinal direction and "ego.y" the one in lateral direction, namely the number of lane changes. "sur.x" represents the average value of number of actions of surrounding vehicles in longitudinal direction and "sur.y" the total number of actions of surrounding vehicles in lateral direction. Under each category, taking "ego.x" as an example, "$\overline{noa}$" is the mean value of number of actions of ego-vehicles in this track. Column "0.9" shows the 90th percentile value and "0.95" the 95th percentile value. The reference value used for the category "ego.x" should cover the majority of the situations, However, for different tracks, the values in column "0.95" are very different. It is noticeable, that these values are proportional to the values of column "$\overline{noa}$" to a certain extent. When dividing the values of column "$\overline{noa}$" with the ones of "0.95", following results are obtained: 63.4, 55.28, 55.28, 56.88, 54.74, which are very close to each other. This value can be understood as the number of frames per action. To ensure that a reference value can cover most cases and yet not to large. A value of 50, which is smaller than the results just calculated is selected for the determination of a reference value (Eq.(3.10)). Namely, for reference value one action is defined for every 50 frames.

$$noa_{ego,x,ref} = \frac{n_{ego,frame}}{50} \tag{3.10}$$

$n_{ego,frame}$ is the number of frames of ego-vehicle. $noa_{ego,x,ref}$ represents the reference value of number of actions in longitudinal direction for ego-vehicle. Larger number of frames leads to larger reference value.

Table 3.2: Number of actions of several tracks

		ego.x			ego.y			sur.x			sur.y		
Nr	$\overline{nof}$	$\overline{noa}$	0.9	0.95	$\overline{noa}$	0.9	0.95	$\overline{noa}$	0.9	0.95	$\overline{noa}$	0.9	0.95
1	317	1.63	4	5	0.13	1	1	1.08	2	2.5	0.49	2	2
7	276	1.55	4	4	0.12	1	1	1.03	2	2.33	0.59	2	2
8	276	1.39	3	4	0.17	1	1	0.86	1.6	2	0.78	2	3
12	455	3.82	7	8	0.07	0	1	2.69	4	4.43	0.64	2	2
25	1040	9.94	16	19	0.06	0	1	4.46	7.05	8.13	0.66	2	2

A surrounding vehicle usually does not stay in the ROI of ego-vehicle through the entire scenario. It can for instance only be present in the first few seconds, then leave the ROI due to slower speed or lane change. Or it can appear in the middle of the scenario due to cut in. Therefore, the number of frames of surrounding vehicle in a scenario (from ego-vehicle's perspective) is not larger than ego-vehicle. This will lead to a smaller number of driving states and smaller number of actions. Due to this reason the reference value used for surrounding vehicles $noa_{sur,x,ref}$ will only be half of ego-vehicles.

$$noa_{sur,x,ref} = \frac{n_{ego,x,ref}}{2} \tag{3.11}$$

As mentioned above, the number of actions of a vehicle in lateral direction corresponds with the times of lane change. $noa_{ego,y}$ and $noa_{sur,y}$ represent the number of actions in lateral direction of ego-vehicle and of all surrounding vehicles respectively. Since lane change is not frequent in a short road section (category "ego.y" and "sur.y" in Table 3.2) and can increase the level of complexity of the scenario once it happens, so the number of actions in lateral direction will not be normalized separately with a reference value. Eq. (3.12) shows the normalized value of number of actions by taking both longitudinal and lateral direction into consideration.

$$noa_{ego,norm} = \left(\frac{n_{ego,x,value}}{n_{ego,x,ref}} + noa_{ego,y}\right)/2 \tag{3.12}$$

$n_{ego,x,value}$ is the number of actions in longitudinal direction. $noa_{ego,norm}$ is the value after normalization. The calculation is similar for surrounding vehicles in the scenario. The only difference is that, for surrounding vehicle, the average value of number of actions in longitudinal direction $\overline{n_{sur,x,value}}$ is used.

$$noa_{sur,norm} = \left(\frac{\overline{n_{sur,x,value}}}{n_{sur,x,ref}} + noa_{sur,y}\right)/2 \tag{3.13}$$

The ten factors from the previous work (with improvement) and the new factor time-to-brake are scene-based and obtain a result for each different scene. The factors number of actions for ego-vehicle and for surrounding vehicles are scenario-based and obtain a result for each scenario, which consists of a series of consecutive scenes. For the calculation of complexity these two factors will be considered the same for all scenes. Thus, 13 factors in total are available for the evaluation of complexity (Table 3.3).

Table 3.3: Summary of all influence factors

Nr.	Influence factor
1	Types of surrounding vehicles within ROI
2	Number of surrounding vehicles
3	Number of connections between traffic participants within ROI
4	Dynamic of surrounding vehicles
5	Variation of dynamical parameters of the surrounding vehicles
6	Deviations of the surrounding vehicles from the predicted trajectories
7	Number of possible actions of ego-vehicle
8	Number of possible actions of surrounding vehicles
9	Time-gap between ego-vehicle and surrounding vehicles
10	Time-to-Brake
11	Blind-spot area
12	Number of actions of ego-vehicle performed in the scenario
13	Number of actions of surrounding vehicles performed in the scenario

In the previous work, the average value of complexities of all scenes contained in a scenario is used as the measure for scenario complexity. Four the accuracy of the result, only the frames with complete ROI are used for the calculations. Frames, in which ROI is partially outside the road section, are cast away. The problem of using the average value is that some complex scenes in a scenario might be ignored due to low level of complexity of the rest of the scenes in this scenario. Regardless of the duration of complex situation, it always has the possibility to turn into a critical situation. To make up for this shortage the maximal complexity of all scenes contained in a scenario is selected in this thesis to indicate the complexity of the scenario (Eq. (3.14)). In this case, the ROI does not need to be completely in the highway section. All frames of a vehicle can be used for the calculation of complexity.

$$\mathrm{C}_{\mathrm{scenario}} = \max\left(\boldsymbol{C_{scene}}\right) \tag{3.14}$$

$\boldsymbol{C_{scene}}$ is a vector containing the complexity of all scenes in the scenario.

3.3 Analysis of influence factors

The major goal of this thesis is to find out more information from the influence factors. In the first part of this subchapter, a sensitivity analysis is conducted to study the relationship between different factors and to offer theoretical support for the determination of relative importance of each influence factor. In the second part a questionnaire is designed for experts who have experience in this field. Their opinions towards these influence factors will be asked. The weighting factors of influence factors will be determined by taking the results of the questionnaire, namely the opinions of experts into consideration.

3.3.1 Sensitivity analysis with equal weighting factors

First it is necessary to clarify some concepts. According to SALTELLI ET AL.'s work [5], sensitivity analysis can be used to determine the relative importance of each input factor. To be more accurate, sensitivity analysis helps to study how the variation of the output is distributed among the variation of inputs. In other words, if varying one input and keeping other inputs fixed, the degree of caused variation at output is different for different inputs. This character can be more accurately defined as the influence level of an input factor. It depends on the distribution of the input factor and is different from the concept “importance degree”. A factor considered to be important by experts, whose variation may not be very influential on the final output. Therefore, in this work it is distinguished between the “influence level” and “importance degree” of an influence factor. The former will be analyzed with sensitivity analysis in this section.

A method for visualization and three indicators is introduced for sensitivity analysis in section 2.2. Scatterplot has the advantage of offering an immediate visual description of the relationship between each influence factor and the complexity compared with using indicators. Compared with derivative, sigma-normalized derivative (SND) and SRC are more robust and reliable. Since sigma-normalized derivative and SRC are equivalent for linear model with enough large amount of data, for simplification the indicator SND will be used for this thesis.

Arbitrary track with large enough number of vehicles is selected for the sensitivity analysis. The scatterplots in the following section are based on the data in track 7 (the 458 vehicles in negative driving direction). Since unlike parameters in physical models, the weighting factors in the linear model used in this thesis cannot be predefined, therefore equal weighting factors will be used for the time being. In section 4.2 the weighting factors will be determined based on the results of questionnaire, a new round of sensitivity analysis will be conducted with the updated weighting factors.

Scatterplots

Track 7 has 458 vehicles, which are 458 scenarios. For the creation of scatterplots, the normalized values of all 13 factors are calculated for each frame of each vehicle in the track, the complexity of each frame is obtained by application of the linear model with equal weighting factors. Since the complexity of a scenario is defined as the maximal complexity of all scenes, the scene with maximal complexity will be extracted and its normalized values and complexity will be used for sensitivity analysis. A scatterplot can be created for each influence factor with its x-axis the normalized value of influence factor and y-axis the value of complexity. Each plot will contain 458 points representing the data of 458 vehicles.

Figure 3.9 shows the scatterplots of all influence factors based on the data in track 7. It is noticeable that for factor “type of surrounding vehicles”, its normalized value concentrates at value 0.5 and 1. Because only two types of vehicles are available in HighD dataset, therefore, as long as there is any surrounding vehicle, the number of types is either 1 or 2, which results in 0.5 and 1 after normalization. For factor “time-to-brake” most points have a normalized factor value 0, only a small amount of points has a value larger than 0. The reason for this phenomenon is that, in most cases a safety distance is kept from the preceding vehicle, the possibility of occurrence of a critical situation is very small. For factor “time-gap” is the same situation. These scatterplots offer an important information, that the distribution of normalized values of different influence factors are different. For some a value larger than 0.5 is normal, while for some others it is rare.

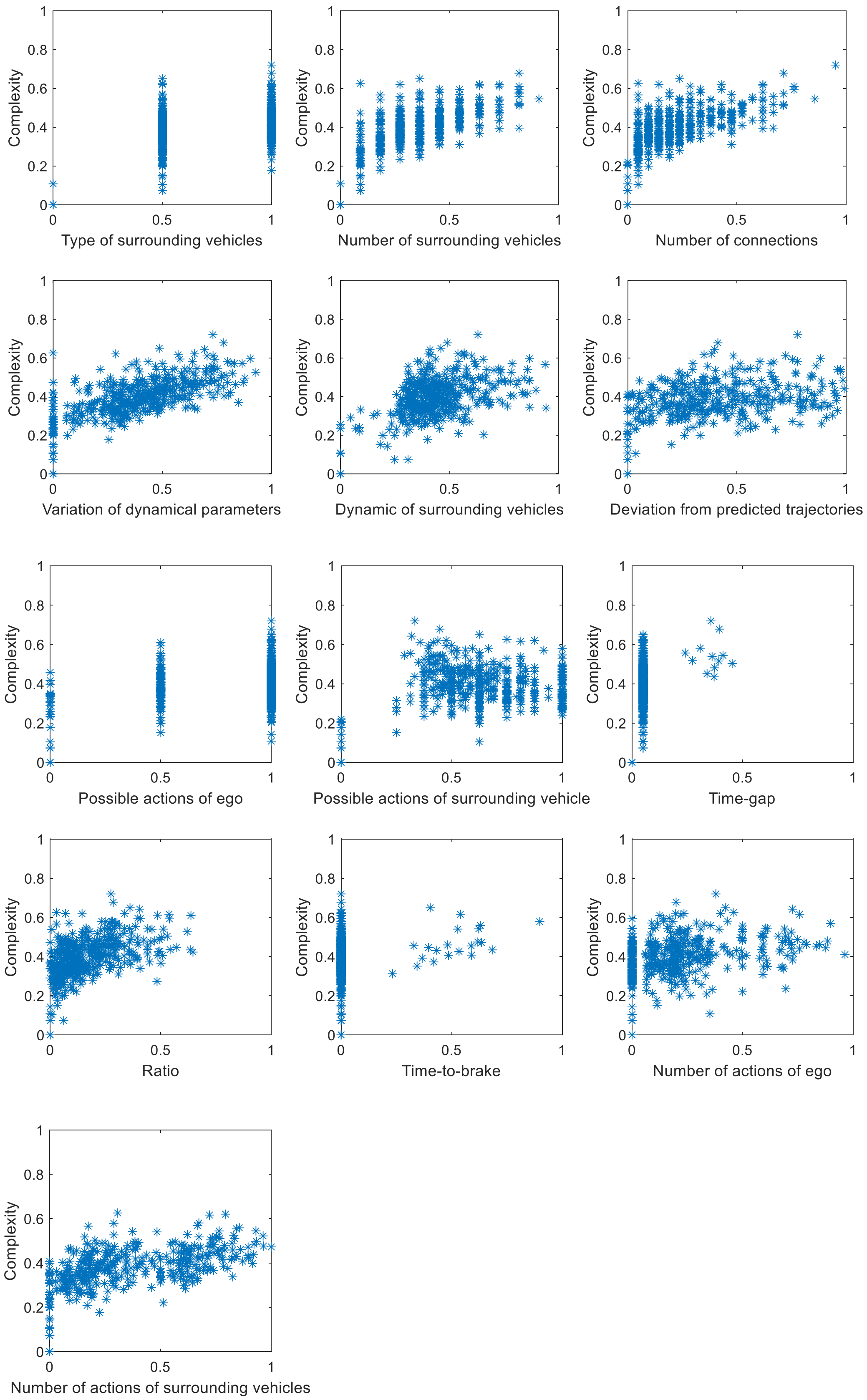


Figure 3.9: Scatterplots of all influence factors for vehicles from the 7th track (only vehicles in negative driving direction)

With help of scatterplots not only the relationship between inputs (respective influence factor) and output (complexity) can be studied, the relationship between different inputs (influence factors) can be checked as well. When comparing each factor with every other factor, 78 scatterplots need to be created for 13 factors, which is very time consuming. For simplification, relationships between factors, which obviously are not related to each other, such as “time-gap” and “ratio”, “possible actions of ego-vehicle” and “number of actions of ego”, etc. will not be included. The scatterplots of several selected pairs are shown in Figure 3.10.

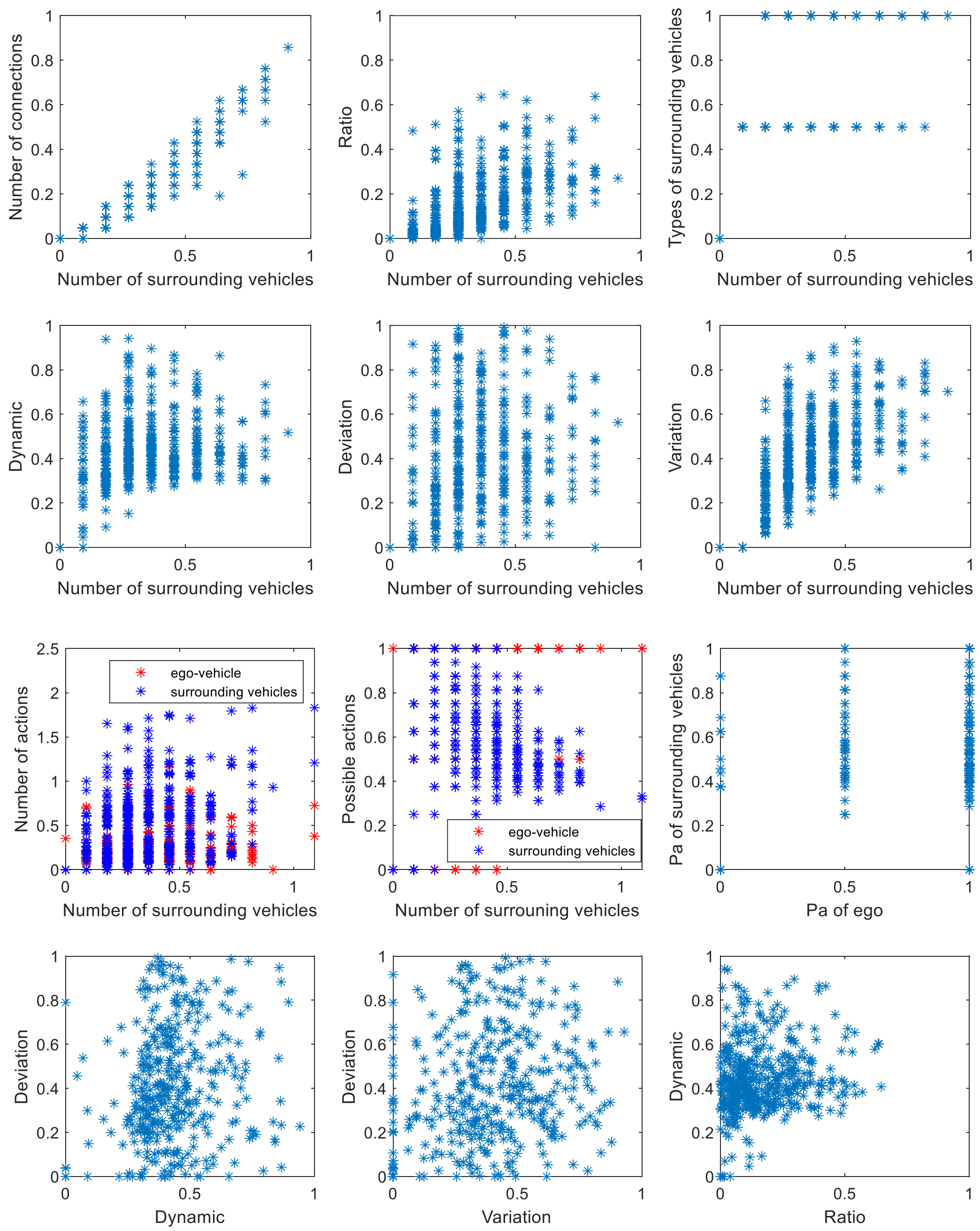


Figure 3.10: Relationships between different influence factors (7th track)

A relative clear linear relationship can be observed between factor “Number of surrounding vehicles” and factor “Number of connections”. As the number of surrounding vehicles increases, number of connections increases as well. Other than this pair of factors, the relationship within other pairs are not very clear. Therefore, it can be concluded, that in general the 13 influence factors are independent from each other, which is a precondition for the following sensitivity analysis.

Although scatterplot offers a good possibility to visualize the influence level of each influence factor on complexity and the relationships between different influence factors. However, this evaluation is only qualitative, the influence level of different factors cannot be compared. In order to find out the most influential and least influential factors, quantitative method is necessary, which is the sensitivity analysis adapted in the following section.

Sigma-normalized derivative

One precondition for the application of sigma-normalized derivative for the measurement of influence level of each input is that, these inputs should be independent from each other, which has already been proved Therefore, sigma-normalized derivative is used to obtain theoretical support for influence level of each factor.

Based on the data used for scatterplots, the normalized values for all influence factors and complexity of each scenario (vehicle) is now available. The standard deviation of each influence factor for all vehicles can easily be calculated with help of MATLAB or EXCEL. The SND can be obtained by using Eq. (2.10). Table 3.4 shows the results of the 7th track. Column "$S_{x_i}^{\sigma}$" contains the value of SND and column "Ranking" is the rank of factors when ordering the influence factors with respect to SND. Factor with the highest ranking (1) has the largest SND and is supposed to be the most influential factor.

Table 3.4: Sensitivity analysis with sigma-normalized derivative for the 7th track

Nr.	Influence factor	$S_{x_i}^{\sigma}$	Ranking
1	Types of surrounding- vehicles	0.2094	4
2	Number of surrounding-vehicles	0.1401	9
3	Number of Connections	0.1392	10
4	Variation	0.1769	6
5	Dynamic	0.1490	8
6	Deviation	0.2790	2
7	Pa of ego-vehicle	0.1505	7
8	Pa of surrounding-vehicles	0.2158	3
9	Time-gap	3.03E-16	13
10	Ratio	0.1364	11
11	Time-to-Brake	0.1201	12
12	Noa of ego-vehicle	0.1930	5
13	Noa of surrounding vehicles	0.3322	1

To avoid accidental results, such analysis is conducted for three more tracks with large enough amount of vehicles, including track 1 (negative driving direction, labeled with "ew"), track 8 and track 15 (positive driving direction, labeled with “we”),. Figure 3.11 shows the ranking of all influence factors of these four tracks, which are highly consistent.

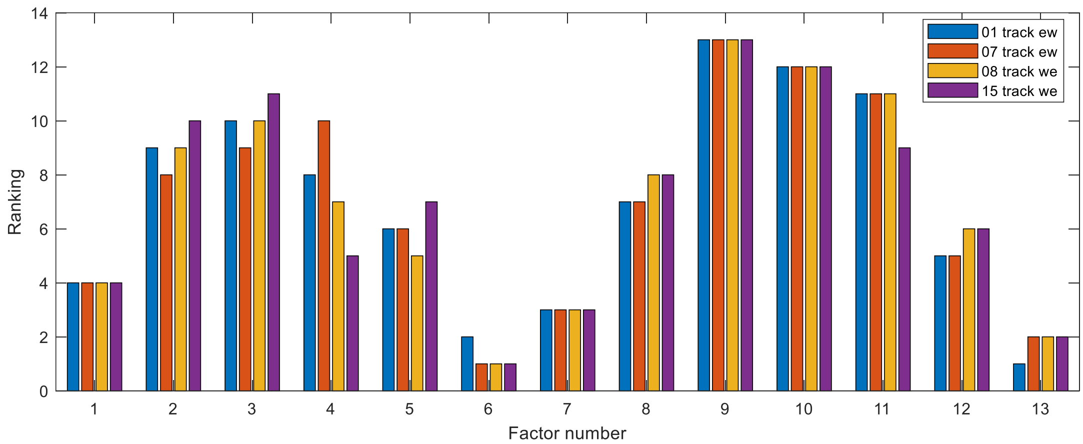


Figure 3.11: Ranking of influence factors with help of sigma-normalized derivative

On average factor 6 “deviation of surrounding vehicles from predicted trajectories” has the highest rank, which means that this factor is most influential on the complexity. Factor 9 has on the contrary the lowest rank and has least influence on the complexity. When looking at further several factors with high ranks, such as factor 1 “types of surrounding vehicles”, factor 7 “possible actions of ego vehicle”, factor 13 “number of actions of surrounding vehicles”, these are the ones with majority of normalized values not less than 0.5, which can be observed in Figure 3.9. On the contrary, factors like “time-gap” and “time-to-brake”, whose normalized values are mostly 0, have very low ranks and have the least influence on the complexity. Factors like “dynamic of surrounding vehicles” and “variation of surrounding vehicles” have most of their normalized values around 0.5, for them it is noticeable that their ranks lie somewhere in the middle as well.

With help of the sensitivity analysis, the influence level of each influence factor can be quantified. However, it does not answer the question of the importance degree of each factor. It is possible, that a very influential factor according to the opinions of experts is not important, or in the opposite, an important factor is not influential on the result. To avoid such situations, it is necessary to rank the influence factors with respect to their importance degree.

3.3.2 Design of the Questionnaire

The ordering of influence factors with respect to their degree of importance will be determined by taking the knowledge and experience of experts into consideration. The opinions of experts will be collected with the help of an online survey conducted in the form of a questionnaire. The questionnaire will be created with the Software Limesurvey [9], which is an open source, web-server-based software. It allows users to design and publish online surveys through web interface. According to SEDLMEIER ET AL. [10, pp. 101-102] creation of a questionnaire usually consists of the following steps, which will be followed during the design process of the questionnaire used for this thesis:

- Theoretical preparation
- Decision on the form of the questionnaire
- Selection of items

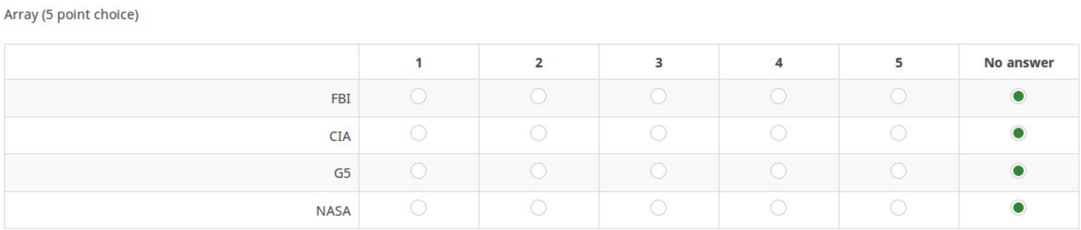
- Analysis of items
- Reliability analysis
- Validity analysis
- Normalization

Theoretical preparation

The goal of designing the questionnaire is to determine the importance degree of the total 13 factors with respect to the assessment of complexity. The content of the questionnaire is based on the study of complexity from previous work [1] and the content from section 3.1 to section 0 in this thesis.

Decision on the form of the questionnaire

The questionnaire is supposed to take the respondents not more than 15 minutes to finish. The attitudes of respondents towards questions like “How important do you consider this factor” or “How complex do you consider this scenario” cannot be quantified. Therefore, the questionnaire will mainly consist of question type in from of array with 3-point choice and 5-point choice. So that the attitudes of respondents can be measured and quantified to a certain extent. Figure 3.12 shows an example of how a 5-point choice array type question looks like in Limesurvey.


Array (5 point choice)

	1	2	3	4	5	No answer
FBI	○	○	○	○	○	●
CIA	○	○	○	○	○	●
G5	○	○	○	○	○	●
NASA	○	○	○	○	○	●

Figure 3.12: 5 point choice – Array type question [11]

Selection of items and analysis of items

The questionnaire of the survey consists of four parts: basic information of respondents, evaluation of influence factors, comparison of factor groups and evaluation of scenario complexity.

In the first part some basic information about the respondents are asked, which include the ages, nationalities of respondents, the type of vehicle they usually drive, driving experience they have, if they work in the field of automobile industry, and if their work is related to safety assessment of automated vehicles. A screenshot of one of these questions is shown in Figure 3.13.

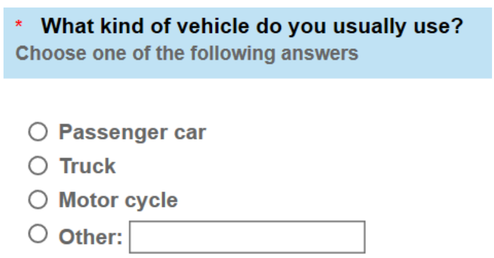


Figure 3.13: Question example in the 1st part of the questionnaire

In the second part of the questionnaire 13 influence factors are required to be evaluated separately (10 factors from the previous work with improvements and three newly added factors introduced in section 3.2). A brief introduction of every factor and an explanation of its eventual influence on the complexity will be given. The question is asked in a form of a 5-point choice. The experts are asked to assign a number between 1 and 5 to the factor indicating its degree of importance with respect to the assessment of traffic situation complexity for automated vehicles (1 corresponds to “not important” and 5 corresponds to “very important”). A comment field is given after evaluation of each factor, so that the opinions of the experts about the factor can be collected, for instance, the reason for their answer or if they have other interpretations of the factor. Figure 3.14 is a screenshot of the first factor as an example.

In the second question group the importance degree of each influence factor is evaluated individually. It is inevitable, that several factors are assigned to the same level of importance. For this situation, the factors need to be compared with each other. However, if a comparison is made between every two different factors, 78 comparisons need to be made in total. This would be very time-consuming for the respondents. To solve this problem, all factors have been divided into the following six groups. Each group will be compared with every other group with respect to their importance degree for the assessment of traffic situation complexity. The comparison will be done with the help of a 3-poin choice array type question. Figure 3.15 shows the matrix to compare Group 1 with all the other groups. 15 such sets of comparison are to be completed. A comment field is offered at the end of this part of the questionnaire, so that the respondents can have the option to express their opinions if they think differently.

- **Group 1 – Dynamic elements**: types of surrounding vehicles, number of surrounding vehicles, number of connections
- **Group 2 – Dynamic**: dynamic of surrounding vehicles, variation of dynamic parameters of surrounding vehicles
- **Group 3 – Possible actions**: possible actions of ego-vehicle, possible actions of surrounding vehicles
- **Group 4 – Number of actions**: number of actions of ego-vehicle, number of actions of surrounding vehicle
- **Group 5 – Precision of actions**: time-gap, time-to-Brake
- **Group 6 – Uncertainty**: deviation of surrounding vehicles from predicted trajectories, blind spot area

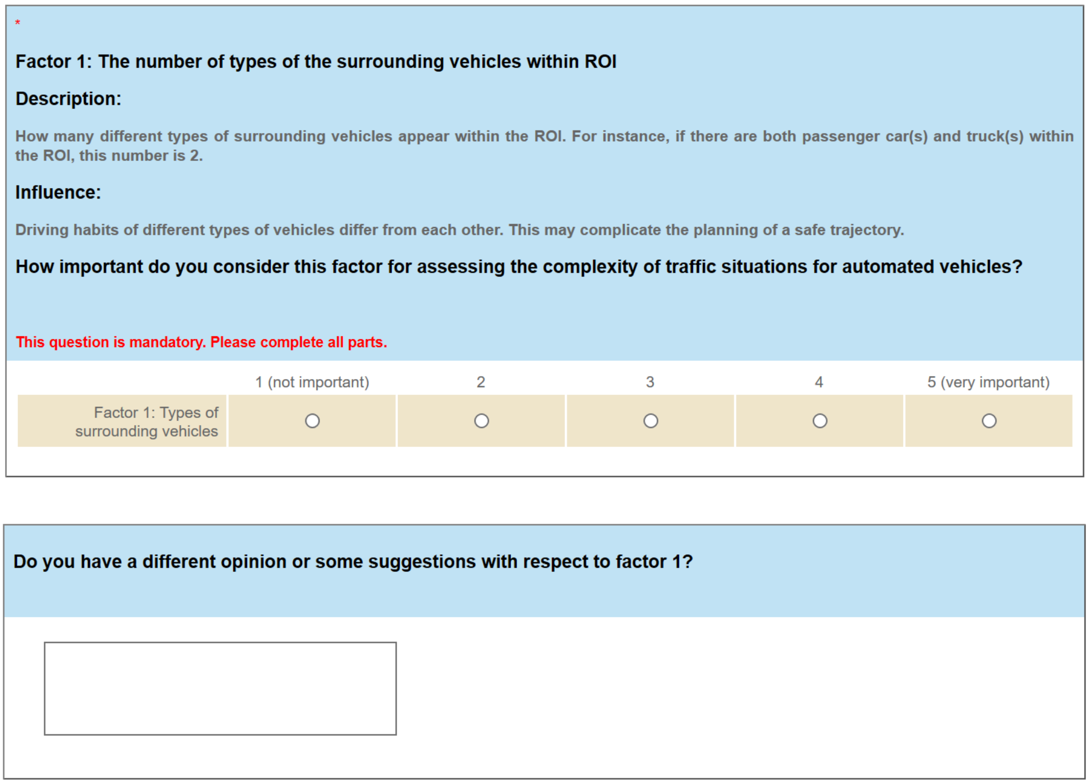


Figure 3.14: Question example in the 2nd part of the questionnaire

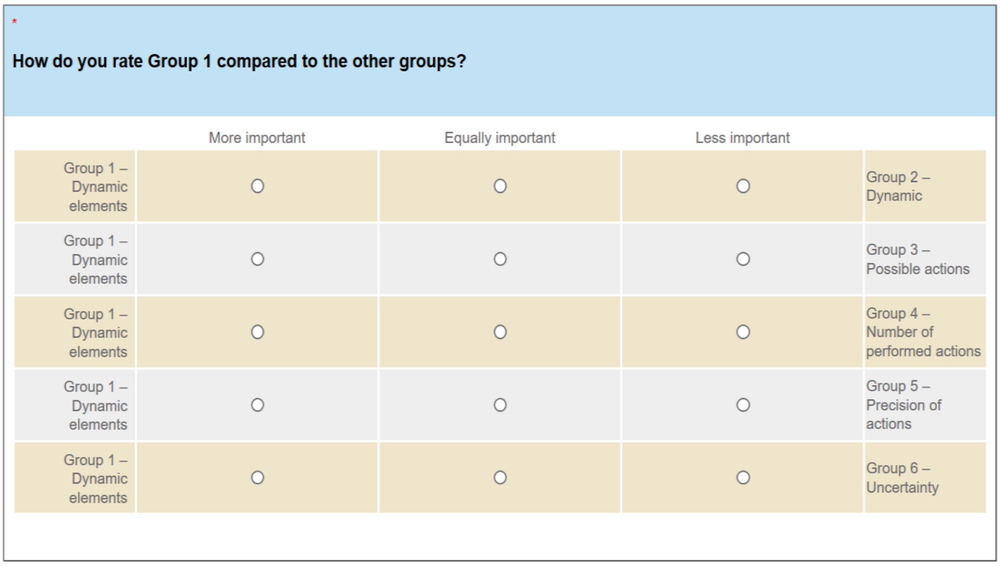


Figure 3.15: Question example in the 3rd part of the questionnaire

Based on the information of the first three parts, the order of influence factors regarding their importance degree and the weighting factors of these factors can already be determined. In the last part of the questionnaire videos of 20 scenarios will be offered. This part of the survey serves mainly as validation for the metric. On the one hand, the complexity of these scenarios will be calculated objectively with the developed metric and the newly determined weighting factors. On the other hand, the respondents are required to evaluate the complexity of each scenario subjectively and assign a number between 1 and 5 to it indicating the complexity level of the scenario

(1 corresponds to “not complex” and 5 corresponds to “very complex”). This question will be created using an array type 5-point choice. The results of the metric will be compared with the answers of respondents. Since the perception and evaluation of the environment is different between sitting in the car and observing the scenario through an animated video, to better support the decision making of respondents, some information will be provided additionally for each scenario. This information cannot be obtained by simply observing the animation video but is fundamental for the evaluation of a scenario. The available information includes the longitudinal velocity of each vehicle in the scenario (Figure 3.16, velocity is marked with yellow next to each vehicle), variation with respect to longitudinal velocity, dynamic of surrounding vehicles with respect to longitudinal velocity and number of actions of surrounding vehicles in longitudinal direction. Not all values of influence factors of a scenario will be offered, so that the respondents will not be affected by the criteria of the developed metric during decision making process.

The selection of the scenarios for the questionnaire is based on the complexity calculated by the metric (13 influence factors with equal weighting factors). So that these scenarios are as representative as possible and cover the variety of the traffic situation, the selection of the scenarios will be based on the following criteria:

- Since in different tracks the traffic flows have different dynamic, the scenarios are selected from several different tracks instead of just one.
- The scenarios should cover as many maneuvers as possible. For instance, cut in, cut out, overtake, follow up, etc.
- The complexities of the scenarios should be distributed as evenly as possible in the range of 0 and 1.
- The additional information provided for all scenarios should cover a range respectively as large as possible.
- Although all vehicles in the scenario are with human drivers, the selected scenarios should fulfil the assumption, that ego-vehicle is an automated vehicle and is supposed to drive intelligently. Behaviors like follow driving at very high speed or overtake from the right side should be avoided.

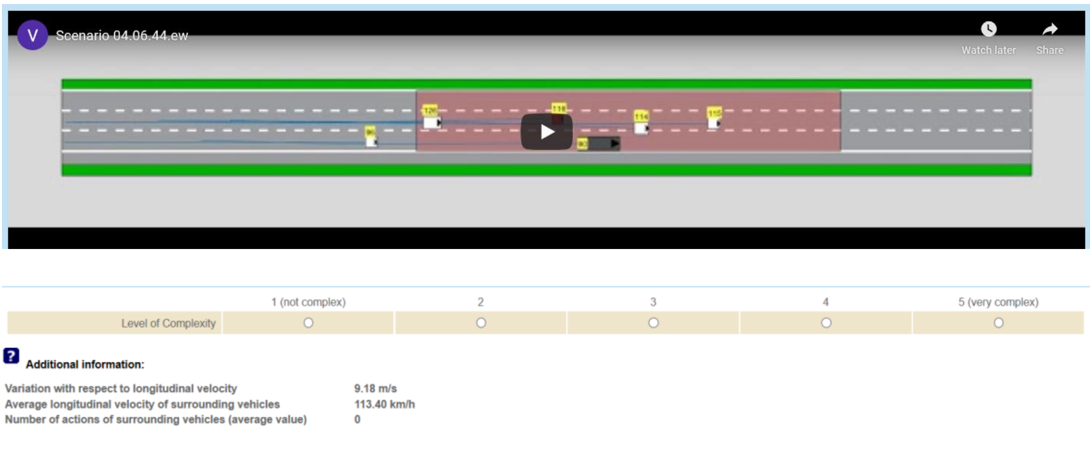


Figure 3.16: Question example in the 4th part of the questionnaire

The selected scenarios are listed in Table 3.6. The information listed from the leftmost column to the rightmost column includes the scenario number, track number (to which track the scenario belongs), ID of ego-vehicle, dynamic with respect to longitudinal velocity ("Dynamic_vx"), variation with respect to longitudinal velocity ("Variation_vx"), average value number of actions of surrounding vehicles ("Noa_sur") and the maximal complexity of all scenes of a scenario ("C_max"). The distribution of scenario complexity ("C_max") and three other values offered as additional information can be seen from Figure 3.18 and Figure 3.21.

Reliability analysis

According to the explanation in SEDLMEIER ET AL.'s [10, pp. 81-82] work reliability indicates the consistency of the test and shows how well a test measures what it should. If a test is reliable means that identical results can be obtained over repeated tests. The test will be unreliable if repeated tests deliver different results.

In the field of social science scale is usually classified into four levels tracing back to STEVENS' [12] work, which are nominal, ordinal, interval and ratio. As introduced above, most of the questions created in this questionnaire are in the form of array with 3-point choice and 5-point choice. Taking the question in Figure 3.14 as an example, the attitude of the respondent towards the importance degree of an influence factor is divided into five levels from not important to very important. The distance between neighboring levels or ranks is unknown and the order of the level is monotone. Such scalar belongs to ordinal scalar. The reliability of a test with ordinal scalar can be quantified with help of a coefficient called Cronbach's Alpha, developed by American educational psychologist CRONBACH [13]. Usually a test will be considered satisfactory in regard to satisfactory when the Cronbach's Alpha is larger than 0.7 [14] (Table 3.5).

Table 3.5: Cronbach's Alpha

Cronbach's Alpha	**Reliability / Intern consistency**
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	poor
$0.5 > \alpha$	Unacceptable

The coefficient α will be calculated with the following formula:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} V_i}{V_t}\right) \tag{3.15}$$

n is the number of items. In the second part of the questionnaire 13 factors are to be evaluated, so the number of items in this case in 13. V_i is the variance of item i and $V_t$is the variance of the total score. Evaluating a respondent the first factor with 5 (very important), then the score for this item will be 5. Total score is the sum of scores of all items.

Validity analysis

A test will be considered valid, when it measures what it claims to measure [10, pp. 85-88]. In the context of this thesis, if the tests measure the complexity of the traffic situation from the perspective of an automated vehicle, or if the selected items really the latent features ought to be measured. A common example is that if a IQ-test truly measures the intelligent level of respondents or does its content actually measure something else like EQ. There are different aspects of validity, the one involved in this work refers to content validity. The evaluation of content validity can be very subjective since it requires the knowledge of experts in relevant fields. Since the topic of this work is relatively new and the development of the metric is based on a lot of research done by people with knowledge and experience in this field. Therefore, the tests in the questionnaires will be for the time being considered as valid.

There is no necessary connection between the reliability and validity of a test. A valid test can be unreliable, and a reliable test can be invalid. Shows a good visualization of the relationship between these two concepts.

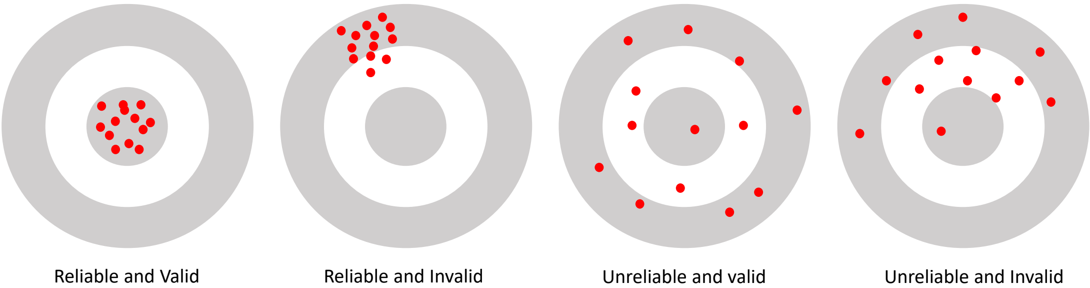


Figure 3.17: Reliability and validity

Normalization

This step is necessary when the respondents are divided into different groups according to some characteristics, which lead to the difference of their answers. For instance, the height ranges between men and women are different. Since the respondents of the questionnaire for this thesis is not divided into different groups, this step is not necessary for this work.

Table 3.6: Selected Scenarios for the questionnaire

Nr.	Track Nr.	Ego-ID	Dynamic_vx	Variation_vx	Noa_sur	C_max
1	04	577	113.4	9.18	0	0.4935
2	04	996	94.84	2.28	0	0.5039
3	05	163	74.1	1.46	0.22	0.4606
4	08	934	111.89	12.63	0.06	0.4354
5	10	265	35.6	0.03	0	0.1678
6	10	760	158.91	9.13	1	0.6214
7	10	820	102.73	1.96	1.5	0.2367
8	11	298	85.86	2.9	1.75	0.3093
9	11	636	96.95	4.53	2.56	0.5227
10	11	1081	75.89	1.86	0.33	0.3273
11	11	1347	115.34	2.69	2	0.3514
12	12	1051	102.27	7.46	0.49	0.894
13	12	1116	71.46	2.72	0.5	0.6482
14	12	1123	63.37	4.91	0.69	0.6061
15	12	1317	60.61	5.62	0.17	0.6823
16	12	1322	60.3	4.82	0.13	0.7726
17	22	31	124.8	16.25	2	0.6703
18	25	371	23.75	3.31	6.25	0.4203
19	25	900	37.58	5.06	4.18	0.5615
20	25	2173	45.82	0.88	4.8	0.383

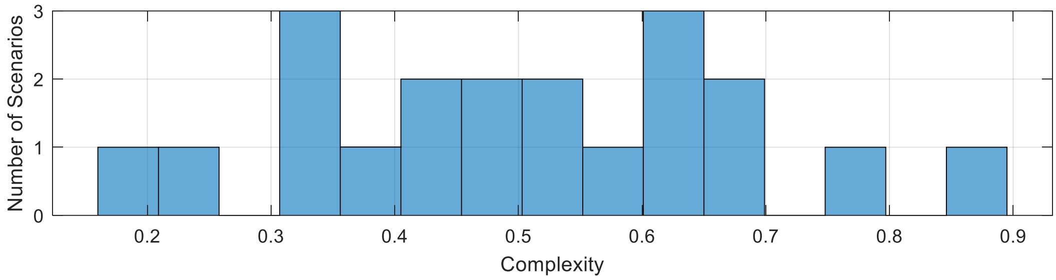


Figure 3.18: Distribution of maximal complexities of selected scenarios

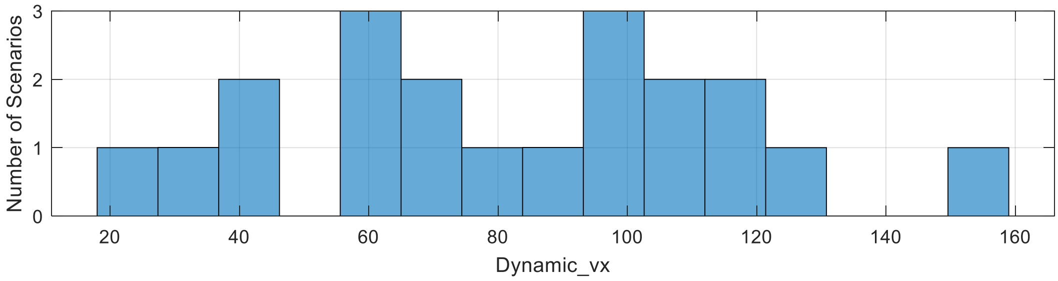


Figure 3.19: Distribution of dynamic with respect to longitudinal velocity of selected scenarios

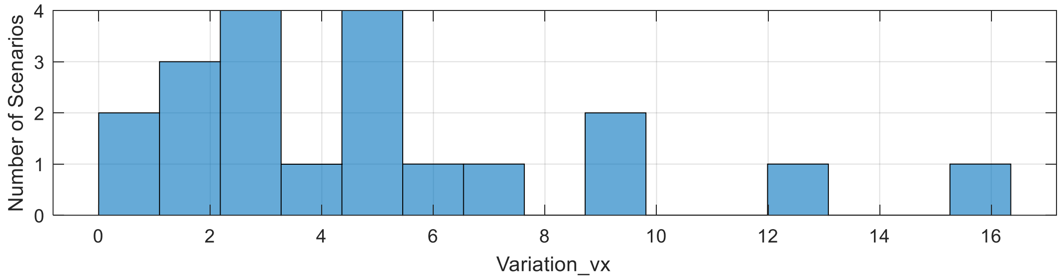


Figure 3.20: Distribution of variation with respect to longitudinal velocity of selected scenarios

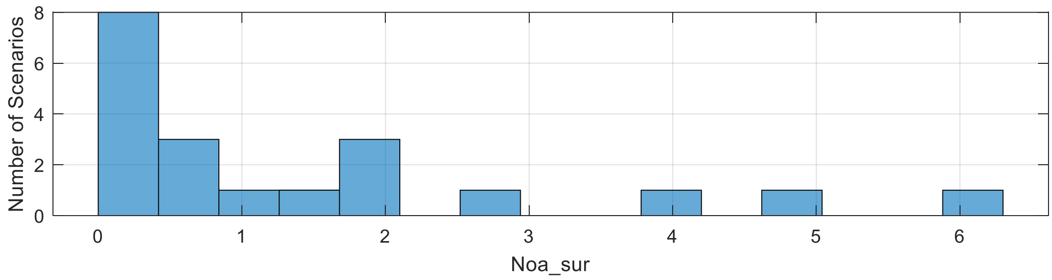


Figure 3.21: Distribution of number of actions of surrounding vehicles of selected scenarios

4 Results

In the first part of this chapter, the results of the questionnaire will be presented and analyzed, the weighting factors of the influence factors will be determined based on these results. In the second part of this chapter, the complexities of 20 scenarios offered in the questionnaire will be calculated with newly determined weighting factors, to compare the results of the develped metric and the opinions of the experts. In the third part, a sensitivitty analysis will be conducted with new weighting factors.

4.1 Results from the questionnaire

This subchapter will present and analyze the responses from the experts. In total 40 responses are collected, 20 of them are completely finished and the rest are only partially answered.

Basic information of respondents

In the first part of the questionnaire some questions about the basic information of the respondents are asked. 35 people have answered all the questions in this part. Amough the 35 people, the majority of them are from Germany. The distribution of nationalities of the respondents is shown in Figure 4.1. The respondents have an average age of about 31 years old. 70.27% of them, namely 26 of them have driving experience larger than 10 years. Except one respondent, the rest 34 respondents have passanger car as their mobile vehicles. 78.38% of the respondents (29) work in the automobile industry and 56.76% (21) have experience in the field of safety assessment of automated vehicles.

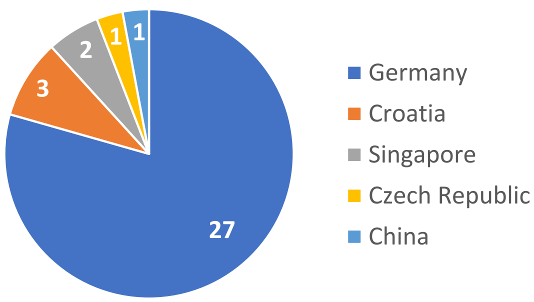


Figure 4.1: Nationalities of the respondents

As can be seen that, the majority of the respondents have many years of driving experience. The majority of them are in the automobile industry and their work is related to the safety assessment of automated vehicles, this makes their responses for the questions asked in the survey valuable and can be used for the research.

Evaluation of influence factors

In the second part of the questionnaire, the respondents are required to evaluate the importance degree of each influence factor independently with respect to the assessment of traffic situation complexity from the perspective of ego-vehicle. 25 respondents have answered all the questions in this part. 20 of them work in the automobile industry and 17 of them have experience in the field of safety assessment of automated vehicles. According to the work of RICHARDSON [15, p. 79], a minimum number of 20 data records can already indicate the validity of the test. Therefore, 25 data records would be sufficient for the following analysis.

The results of this part are shown in Table 4.2. The indices from 1 to 13 in the first row indicate the first to the thirteenth factor (the factor corresponding to each index see Table 3.3). The indices from 1 to 25 in the leftmost column represents the IDs of the 25 respondents. Starting from the second row, each row represents a respondent's assessment of the importance degree of the 13 factors (assessing with a number from 1 to 5 indicating ascending importance degree, from "not important" to "very important"). For instance, the element in the third row and the fifth column means that, the respondent with ID 2 consider the fifth factor as very important. The evaluations of all factors by one respondent adds up to the result in the rightmost column under the name "sum", which is also called "total score" in statistics.

Reliability analysis

First step is to analyze the reliability of the test, namely how well this part of the questionnaire measures the importance degree of the factors or if the responses are internally consistent. Table 4.1 lists the variance of each influence factor V_i $(i = 1,2, \dots ,13)$ and the variance of the total score V_t. With help of the formula Eq.(3.15) introduced in section 3.3.2, Cronbach's Alpha for this test can be calculated as follow:

$$\alpha = \frac{13}{13-1}\left(1 - \frac{\sum_{i=1}^{13} V_i}{V_t}\right) = 0.8216 \tag{4.1}$$

α has a value between 0.8 and 0.9. According to the criteria in Table 3.5 the test is evaluated as "good" with respect to reliability and the data records can be used for further analysis.

Table 4.1: Variance of influence factors

V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_t
1.24	0.91	1.06	1.00	1.50	0.88	1.51	0.99	1.76	1.59	1.04	1.67	1.42	68.51

Kruskal-Wallis test

The purpose of this part of the questionnaire is to find out the importance degree of each influence factor and try to rank these factors according to their importance degrees. First step is to find out if these factors really differ from each other with respect to importance degree. The data records in Table 4.2 have the following characteristics: first, the items (factors) are independent from each other. Second, ordinal scales are used for the evaluations. Third, the distribution of evaluations of each item does not necessarily conform to a normal distribution or have specified parameters. The method used to study the central tendencies of such data records is the Kruskal-Wallis test, also known as "H test", developed by KRUSKAL ET AL. [16, pp. 585-587] in 1952. The advantage of this method is that, it has very low requirements for the distribution of data and the items do not necessarily to have equal number of samples (in case of Table 4.2 all items have equal number of samples 25). This method calculates the mean rank of each item and tests if these mean ranks differ from each other. In the field of statistics, a null hypothesis (denote H_0) is proposed before testing. It will be assumed that there is no association among the data until the results overrule this assumption. In this case, the null hypothesis is that, there is no difference between the influence factors with respect to importance degree.

Table 4.2: Evaluation of the influence factors

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**sum**
1	1	5	3	4	3	5	4	4	2	2	5	2	2	42
2	3	4	3	5	5	5	5	4	2	5	5	3	5	54
3	2	4	4	4	3	4	3	4	2	2	3	2	3	40
4	2	5	5	4	1	5	5	5	5	5	5	1	1	49
5	3	3	4	4	4	3	3	3	3	3	2	3	3	41
6	3	2	4	4	3	2	2	2	4	4	4	3	3	40
7	2	4	4	4	5	4	3	4	5	4	3	4	4	50
8	3	4	2	3	4	5	2	4	2	2	4	3	4	42
9	2	4	3	5	5	5	5	4	2	5	5	4	5	54
10	2	4	4	5	5	5	5	4	2	5	5	4	5	55
11	5	5	5	5	5	5	5	5	5	5	5	5	5	65
12	4	2	3	4	5	4	2	2	2	1	3	1	5	38
13	3	4	4	5	1	5	4	2	5	2	3	1	5	44
14	4	3	1	1	5	5	1	3	3	3	5	4	4	42
15	4	4	4	5	5	5	3	3	5	5	5	3	3	54
16	4	4	5	5	5	4	3	4	5	3	4	4	4	54
17	2	3	3	4	3	4	3	2	4	4	3	2	2	39
18	5	5	5	5	5	4	5	5	5	5	5	4	4	62
19	4	5	5	5	4	5	5	5	4	4	4	5	5	60
20	4	5	5	4	4	4	4	3	4	4	3	4	3	51
21	5	5	5	5	5	4	4	4	5	5	5	4	4	60
22	2	4	3	2	4	5	5	5	3	3	5	5	5	51
23	3	2	4	4	3	3	2	4	1	2	2	2	4	36
24	3	4	4	5	5	2	3	3	3	3	4	5	5	49
25	2	4	4	4	3	5	3	3	3	3	4	2	2	42

The mean rank of each item will be calculated in the following way. All elements in Table 4.2 (except the indices and the rightmost column) are ranked in an ascending order (from 1 to 5) regardless of the items they belong to. After ordering these 325 (25 x 13) elements will eventually look like 1, 1, …, 1, 2, 2, …,2, 3, …, 3, 4, …, 4, 5, …, 5 and have a rank from 1 to 325. The ranks of elements belonging to the same item will be added and the rank sum of each item can be obtained. Mean rank of each item is obtained by dividing its rank sum by the number of elements owned by it, which is 25 for all items. The results are shown in Table 4.3. The rank sum for some items is not a integer but a decimal number. The reason for this is that the tied elements, e.g. the elements with the same rank share the average rank. For instance, two elements with the same value tied for the fourth and fifth rank will share the average rank of 4.5. A visualization of the mean rank of all influence factors is shown in Figure 4.2. As can be seen that, factor 6 “deviation of the surrounding vehicles from the predicted trajectories” has the highest mean rank and factor 1 “types of surrounding vehicles” has the lowest mean rank.

Table 4.3: Mean rank of each influence factor

Item/Factor	Number of elements	Rank sum	Mean rank
1	25	2766.50	110.66
2	25	4342	173.68
3	25	4207	168.28
4	25	5016	200.64
5	25	4683.50	187.34
6	25	5165.50	206.62
7	25	3736.50	149.46
8	25	3760.50	150.42
9	25	3570	142.80
10	25	3762	150.48
11	25	4639.50	185.58
12	25	3109	124.36
13	25	4217	168.68

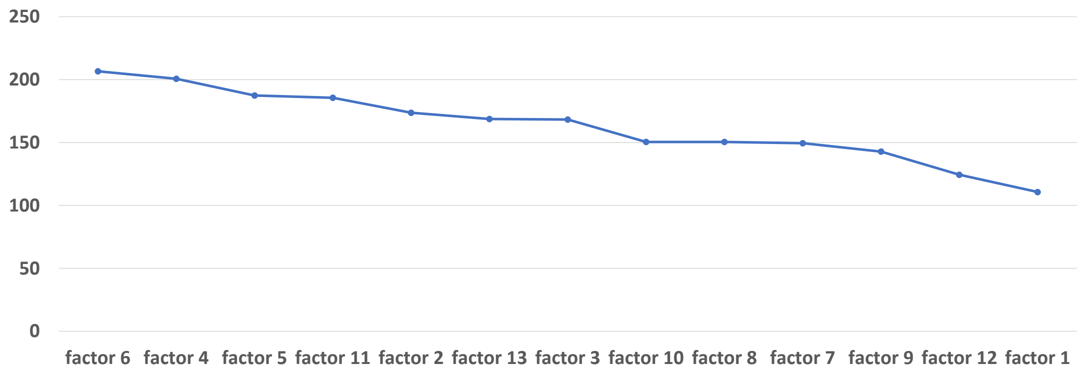


Figure 4.2: Mean rank of all influence factors

With the mean rank of each item, the test statistic of this method H can then be calculated with the following formula, it indicates the discrepancies of the rank sums of different items:

$$\mathrm{H} = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N-1) = 29.668, \qquad N = \sum_{i=1}^{k} n_i \tag{4.2}$$

n_i is the number of elements of item i, in this case they are all equal to 25. N is then the total number of elements 325. k is the total number of items which is 13. R_i represent the rank sum for item i. Add all these values to the formula and obtains 29.688 for the test statistic. With this result the significance, denoted p-value, can be examined. Significance of the result is defined as the probability that an extreme result can be obtained if the null hypothesis holds [10, pp. 371-401]. Usually with a significance level $p < 0.05$ the null hypothesis can be rejected [17, p. 43] or the corresponding critical value is smaller than the value of H for $p = 0.05$. The number of items in this test is larger than 5 and therefore H satisfies the condition for the application of Chi-square distribution, also known as χ^2-distribution [10, pp. 549-550] (Chi-square distribution does not work well for small size data). Part of the Chi-square distribution table is shown in Figure 4.3. The horizontal indices represent the significance level p and the vertical indices under the name “df” the degree of freedom of the test obtained by the number of items minus one:

$$\mathrm{df} = \mathrm{k} - 1 \tag{4.3}$$

In this case df has the value 12 ($\mathrm{k} = 13$). As can be seen from the table that, with the significance level $p = 0.05$ and the degree of freedom 12, the critical value of 21.03 can be extracted. If the value of test statistic H is larger than is critical value, the difference will be considered as significant and the null hypothesis can be rejected. The H calculated by Eq. (4.2) is larger than the critical value (29.668 > 21.03). Therefore, the assumption that all influence factors having same importance degree fails. The central tendencies of these factors do differ from each other.

	p								
df	**0.995**	**0.99**	**0.975**	**0.95**	**0.9**	**0.1**	**0.05**	**0.025**	**0.01**

11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
…									

Figure 4.3: Part of Chi-square distribution table [18]

Dunn’s test

Until now with help of the Kruskal-Wallis test, a conclusion can be made, that the 13 influence factors are not equally important with respect to the assessment of traffic situation complexity from the perspective of ego-vehicle. However, the conclusion does not answer the question, which factor differs from which factor and how significantly. To solve this problem a post-hoc-test is necessary. The most used method for such problems is the Dunn’t test [19], with which multiple pairwise comparisons between items can be conducted using rank sums. For k items in total $k(k-1)/2$ pairwise comparisons are to be performed, the number in case of 13 items is 78. The null hypothesis for each pairwise comparison of the test is that, the probability of observing a value randomly from the first group larger than that from the second group is 50%. In other words, there is no significant difference between the two groups or the two items. The conduction of Dunn’s test is based on mean rank obtained in Kruskal-Wallis test. The null hypothesis can be rejected when the following inequality is satisfied [20]:

$$Q_{0.05} < \frac{\left|\bar{R}_i - \bar{R}_j\right|}{\sqrt{\left(\frac{N(N+1)}{12} - \frac{\sum_{s=1}^{r} \tau_s^3 - \tau_s}{12(N-1)}\right)\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{4.4}$$

$\bar{R}_i$ is the mean rank of item i and $\bar{R}_j$ is the mean rank of item j. As already mentioned above, N is the total number of observations, which is 325. n_i and n_j are the number of elements of item i and j respectively. τ_s is the number of tied observations. The summation term will be zero if there is no tied observations. $Q_{0.05}$ ist he critical value for a significance level $p = 0.05$. Difference of mean ranks between every two factors are listed in Table 4.4. The indices in the leftmost column represent the factors sorted by their mean ranks in a descending order. The horizontal indices represent the difference of mean rank between factor i and the factor specified by the number (factor i has a higher mean rank). For instance, the column $i - 8$ shows the mean rank difference between factor i and factor 8 ($i = 6{,}4{,}5{,}11{,}2{,}13{,}3{,}10$). Dunn's test answers the question, that how large this difference is so that the compared factors can be considered significantly different.

Table 4.4: Mean rank difference between items

i	i - 1	i - 12	i - 9	i - 7	i - 8	i - 10	i - 3	i - 13	i - 2	i - 11	i - 5	i - 4
6	95.96	82.26	63.82	57.16	56.20	56.14	38.34	37.94	32.94	21.04	19.28	5.98
4	89.98	76.28	57.84	51.18	50.22	50.16	32.36	31.96	26.96	15.06	13.30	
5	76.68	62.98	44.54	37.88	36.92	36.86	19.06	18.66	13.66	1.76		
11	74.92	61.22	42.78	36.12	35.16	35.10	17.30	16.90	11.90			
2	63.02	49.32	30.88	24.22	23.26	23.20	5.40	5.00				
13	58.02	44.32	25.88	19.22	18.26	18.20	0.40					
3	57.62	43.92	25.48	18.82	17.86	17.80						
10	39.82	26.12	7.68	1.02	0.06							
8	39.76	26.06	7.62	0.96								
7	38.8	25.10	6.66									
9	32.14	18.44										
12	13.7											
1												

There are many programs which can perform this test. The program used in this thesis is the MATLAB function developed by CARDILLO [21]. The comparison starts from the factor with the highest mean rank, namely factor 6. It will be compared with the factor with the second, third, … and at last the smallest mean rank. Then the factor with the second largest mean rank will be compared with the factors with smaller mean ranks in their descending order. Such comparison goes on until the last two factors with smallest mean ranks. The result of Dunn's test is shown in Table 4.5. In total 78 comparisons are conducted. The column with name "Comparison" contains the indices of factors compared with each other. For instance, "6-1" means the comparison between factor 6 and factor 1. Q is the value calculated by the expression on the right side of Eq. (4.4). Column "$Q_{0.05}$" contains the corresponding critical value at significance level $p = 0.05$. The column H_0 indicate if the null hypothesis holds. With "N" the null hypothesis is rejected, which means that there is significant difference between the two factors. With "Y" the rejection of null hypothesis is failed and the difference between two factors is considered as not significant. As can be seen from the result that. Only two pairs of factors are judged to be different from each other significantly, namely factor 6 and 1, factor 4 and 1. When looking at Table 4.4, the difference of mean rank of these two pairs are the two largest differences in the left upper corner of the table. When observing Figure 4.2, factor 6 and 4 are the two factors with the highest mean ranks and factor 1 has the smallest mean rank.

Factor 6 "deviation of the surrounding vehicles from the predicted trajectories" is related to the uncertainty of a scenario. If the behavior of a surrounding vehicle is very different from the result of prediction of ego-vehicle, the risk of getting involved in a critical condition will increase. Factor 1 "types of surrounding vehicles" is considered as the least important factor by the respondents. This factor is included for the assessment for the complexity in the first place is that, different types of traffic participants move differently. For instance, compared with trucks, the movement of passenger cars are more flexible and more dynamic. Several respondents have offered the following arguments based on their experience for their choices: the types of traffic participants can be identified, and the mode of each type can be preprogrammed, so that the behaviors of different types of vehicles can better be predicted and therefore this factor will not have much influence on the complexity.

Comparison between factor groups

In case that the importance degree cannot be distinguished from each other very clearly, which is the situation based on the results of the analysis of the second part of the questionnaire. In the third part of the questionnaire, 13 factors are divided into 6 groups according to the different aspects reflected by them in a scenario. The respondents are required to compare these groups with respect to their importance degree. In total 24 respondents have finished all the questions in this part.

The results are visualized with help of a bar graph in Figure 4.4. The information contained in each bar shows the distribution of attitudes of respondents towards a pair of comparison between two groups illustrated with the label in horizontal axis. For instance, the label of the first bar "Group 1 - 2" means that Group 1 being compared with Group 2. The numbers in the bar show that, 12 respondents think factors in Group 1 are less important than the ones in Group 2, 9 respondents think they are equally important and the rest 3 think that the factors in Group 1 are more important. By observing all bars in the graph, Group 6 is considered to be more important when comparing with any other group and therefore can be seen as the most important factor group. By observing bar "Group 1 - 2 ", "Group 2 - 3", "Group 2 - 4" and "Group 2 - 5", the number of respondents who consider factors from Group 2 more important is obviously larger than the number of respondents with reversed opinions. Therefore, Group 2 can be treated as the second most important factor group. Similarly, Group 5 can be seen as the third most important factor group when observing bar "Group 1 -5", "Group 3 - 5" and "Group 4 - 5". Among the remaining three groups, the number of respondents who think factors from Group 3 are more important, is larger than the number of respondents with reversed opinions. Group 3 be the fourth most important factor Group. By observing bar "Group 1 - 4", the number of respondents with opposite positions are almost the same and the majority respondents think these two groups are equally important. Therefore, Group 1 and Group 4 are judged as equally important.

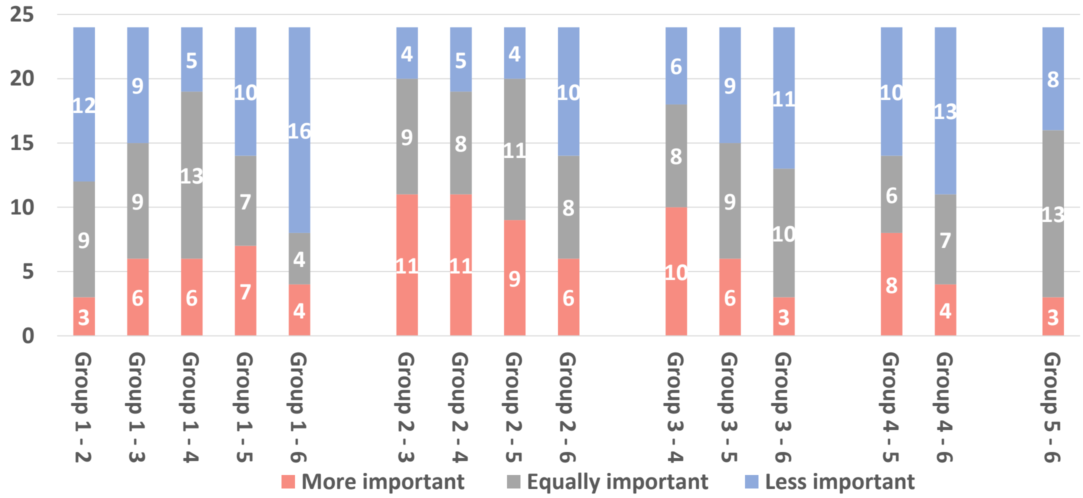


Figure 4.4: Comparison between different factor groups

Based on the analysis above, the groups can be ranked as follow in a descending order according to their importance degree: Group 6 > Group 2 > Group 5 > Group 3 > Group 1 = Group 4. Figure 4.5 shows the result when marking the factor in Figure 4.2 with the group number which they belong to (G6 means Group 6). As can be seen that the four factors with the highest mean ranks belong to the two groups which are considered by the respondents as the most important. The mean ranks of factors from Group 3 and Group 6 lie somewhere in the middle and are not necessarily higher than the ones from Group 1 and Group 4. By observing the bar “Group 1 - 5” and “Group 3 - 5” it can be noticed that, the difference of number of respondents with opposite positions is only 3, which is not very large. This difference is only 2 when comparing Group 4 and 5. In other words, the difference between Group 1 and 4 and Group 3 and 5 is not very large with respect to their importance degree. This does not conflict with the results from the second part of the questionnaire. The order of these groups regarding their importance can now be updated: Group 2 and 6 > Group 3 and 5 ≈ Group 1 and 4.

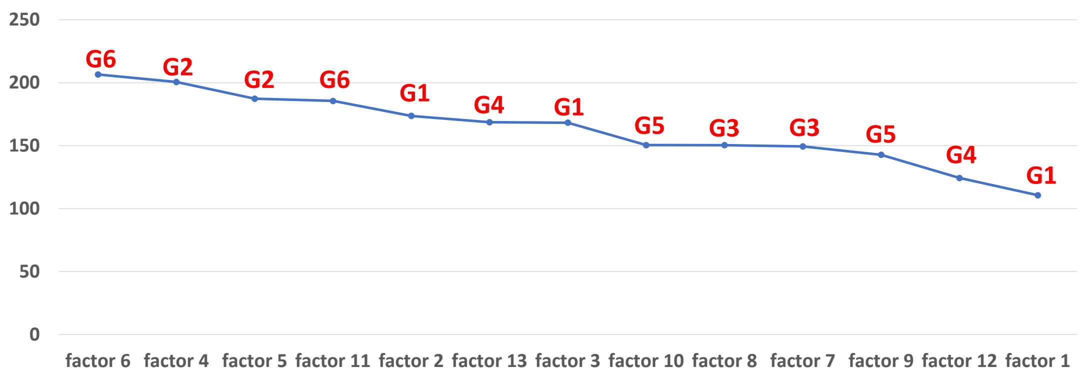


Figure 4.5: Mean rank of all factors with group number

Table 4.5: Results of Dunn's test

Comparison	Q	$Q_{0.05}$	H_0	Comparison	Q	$Q_{0.05}$	H_0
6-1	3.747171	3.40676	N	11-3	0.675553	3.40676	Y
6-12	3.212196	3.40676	Y	11-13	0.659933	3.40676	Y
6-9	2.492127	3.40676	Y	11-2	0.464687	3.40676	Y
6-7	2.232058	3.40676	Y	2-1	2.460887	3.40676	Y
6-8	2.194571	3.40676	Y	2-12	1.925912	3.40676	Y
6-10	2.192228	3.40676	Y	2-9	1.205843	3.40676	Y
6-3	1.49715	3.40676	Y	2-7	0.945774	3.40676	Y
6-13	1.481531	3.40676	Y	2-8	0.908287	3.40676	Y
6-2	1.286284	3.40676	Y	2-10	0.905944	3.40676	Y
6-11	0.821597	3.40676	Y	2-3	0.210866	3.40676	Y
6-5	0.752871	3.40676	Y	2-13	0.195247	3.40676	Y
6-4	0.233515	3.40676	Y	13-1	2.265641	3.40676	Y
4-1	3.513656	3.40676	N	13-12	1.730665	3.40676	Y
4-12	2.978681	3.40676	Y	13-9	1.010596	3.40676	Y
4-9	2.258612	3.40676	Y	13-7	0.750528	3.40676	Y
4-7	1.998543	3.40676	Y	13-8	0.71304	3.40676	Y
4-8	1.961056	3.40676	Y	13-10	0.710697	3.40676	Y
4-10	1.958713	3.40676	Y	13-3	0.01562	3.40676	Y
4-3	1.263636	3.40676	Y	3-1	2.250021	3.40676	Y
4-13	1.248016	3.40676	Y	3-12	1.715045	3.40676	Y
4-2	1.052769	3.40676	Y	3-9	0.994976	3.40676	Y
4-11	0.588083	3.40676	Y	3-7	0.734908	3.40676	Y
4-5	0.519356	3.40676	Y	3-8	0.697421	3.40676	Y
5-1	2.994301	3.40676	Y	3-10	0.695078	3.40676	Y
5-12	2.459325	3.40676	Y	10-1	1.554943	3.40676	Y
5-9	1.739256	3.40676	Y	10-12	1.019968	3.40676	Y
5-7	1.479188	3.40676	Y	10-9	0.299899	3.40676	Y
5-8	1.4417	3.40676	Y	10-7	0.03983	3.40676	Y
5-10	1.439357	3.40676	Y	10-8	0.002343	3.40676	Y
5-3	0.74428	3.40676	Y	8-1	1.5526	3.40676	Y
5-13	0.72866	3.40676	Y	8-12	1.017625	3.40676	Y
5-2	0.533414	3.40676	Y	8-9	0.297556	3.40676	Y
5-11	0.068727	3.40676	Y	8-7	0.037487	3.40676	Y
11-1	2.925574	3.40676	Y	7-1	1.515113	3.40676	Y
11-12	2.390598	3.40676	Y	7-12	0.980138	3.40676	Y
11-9	1.670529	3.40676	Y	7-9	0.260068	3.40676	Y
11-7	1.410461	3.40676	Y	9-1	1.255045	3.40676	Y
11-8	1.372974	3.40676	Y	9-12	0.720069	3.40676	Y
11-10	1.370631	3.40676	Y	12-1	0.534975	3.40676	Y

Evaluation of complexities of scenarios

In the last part of the questionnaire, the respondents are offered the videos of 20 scenarios and are required to evaluate the complex degree of these scenarios subjectively. The questions are asked in the form of 5-point choice. An integer from 1 to 5 is assigned by respondents to each scenario indicating its complexity level, with 1 corresponding to “not complex” and 5 “very complex”. 20 respondents have finished this part of the questionnaire completely and this number is sufficient for the evaluation of the test. All videos are available on the internet and are reachable with the websites listed in Appendix A.

The data type of the results is the same as that in the second part of the questionnaire, which is ordinal scale. Therefore, the methods used in the previous part can be adopted for this part as well. Goal is to rank these scenarios according to their complexity level. First step is to check the reliability of the test. If the test is proved reliable, the Kruskal-Wallis test will then be conducted to see if there is any significant difference among the scenarios regarding their complexity level. If true a post-hoc Dunn’s test will be conducted to discover which scenarios are significantly more complex than the others. Due to larger size of data, the detailed results of the responses cannot be covered in the text. A summarization is shown in Table 4.6. The indices in the leftmost column represent the complexity level. The horizontal indices from 1 to 20 represent the 1st to the 20th scenario. Each column shows the distribution of evaluation by respondents. For instance, for the 5th scenario, 16 respondents have this with 1 (“not complex”) evaluated and 4 respondents with 2.

Table 4.6: Results of the fourth part of the questionnaire

C level	1	2	3	4	5	6	7	8	9	10
1	7	2	4	1	16	15	9	10	10	5
2	6	8	7	10	4	5	5	8	7	14
3	6	7	4	5	0	0	5	2	1	1
4	1	2	3	4	0	0	1	0	2	0
5	0	1	2	0	0	0	0	0	0	0

C level	11	12	13	14	15	16	17	18	19	20
1	4	1	2	0	0	1	2	9	0	2
2	6	8	4	4	2	7	12	8	5	8
3	6	7	5	5	5	8	6	1	8	6
4	4	3	7	6	9	3	0	2	5	4
5	0	1	2	5	4	1	0	0	2	0

Same as in the second part of the questionnaire, Eq. (3.15) is applied for the calculation of Cronbach’s Alpha. In this case, the number of items $n = 20$, V_i is the variance of each item (scenario, Eq. (4.5)). The result is 0.8560 and lies in the interval of 0.8 and 0.9. According to Table 3.5, the test can be evaluated as “Good” with respect to reliability.

$$\alpha = \frac{20}{20-1}\left(1 - \frac{\sum_{i=1}^{20} V_i}{V_t}\right) = 0.8560 \tag{4.5}$$

For Kruskal-Wallis test, null hypothesis H_0 for the test is: there is no difference between scenarios regarding their complex level, all scenarios are assumed to be equally complex. The mean rank of each scenario is calculated, and the results are shown in Table 4.7. Figure 4.6 shows the mean rank of all scenarios is a descending order ("s" in horizontal axis means "scenario"). In the second part, the test statistic H is calculated, then its value is compared with the critical value in Chi-square distribution table for the significance level $p = 0.05$. The null hypothesis can be rejected if H is larger than the critical value. Alternatively, p value can be calculated based on the mean ranks of all items and the result will be compared with 0.05. If p is smaller than 0.05, then there is significant difference among items and the null hypothesis will be rejected. This calculation can be performed with a MATLAB function $kruskalwallis$. After application of this function p has a value of 1.3833×10^{-23}, which is far smaller than 0.05. Therefore, significant difference exists between scenarios regarding their complexity levels. In the next step, Dunn's test is conducted to identify the scenarios which are most significant complex.

Table 4.7: Mean rank of all scenarios

Scenario	N	Rank sum	Mean rank	Scenario	N	Rank sum	Mean rank
1	20	3420	171	11	20	4332	216.6
2	20	4535	226.75	12	20	4839	241.95
3	20	4365	218.25	13	20	5456	272.8
4	20	4576	228.8	14	20	6175	308.75
5	20	1486	74.3	15	20	6508	325.4
6	20	1605	80.25	16	20	4952	247.6
7	20	3069	153.45	17	20	3830	191.5
8	20	2426	121.3	18	20	2802	140.1
9	20	2683	134.15	19	20	5663	283.15
10	20	2908	145.4	20	20	4570	228.5

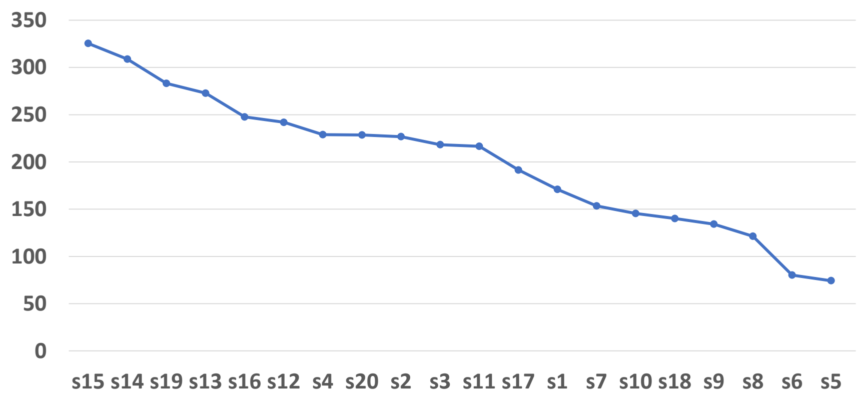


Figure 4.6: Mean rank of all scenarios in a descending order

In the Dunn's test, pairwise comparison is conducted between every two different item (scenario), in this case is 190 pairs of comparison (number of items $k = 20$). The null hypothesis for each comparison is that the involved two scenarios are equally complex. Same calculation as the second part of the questionnaire is conducted. Due to large amount of data, Table 4.8 only shows the pairs, whose null hypothesis are rejected. In other words, there is significant difference between the pair of scenarios regarding their complexity level. As can be seen that the scenarios on the left side of symbol "-" are the ones with higher mean ranks. There is significant difference between these scenarios and the 5th and 6th scenario, which are the two scenarios with the lowest mean ranks. When ordering the scenarios with descending mean rank like in Figure 4.6, from the 11th scenario to the ones with higher mean rank, the difference between this scenario and the 5th or 6th scenario is judged as significant.

Table 4.8: Pairs with significant difference regarding complexity level

Pairs with significant difference

15-5	14-5	19-5	13-5	16-5	12-5	4-5	20-5	2-5	3-5	11-5
15-6	14-6	19-6	13-6	16-6	12-6	4-6	20-6	2-6	3-6	11-6
15-8	14-8	19-8	13-8							
15-9	14-9	19-9	13-9							
15-18	14-18	19-18	13-18							
15-10	14-10	19-10								
15-7	14-7	19-7								
15-1	14-1									
15-17										

Results of the two methods can now be compared. For ordinal scale, median is often used to examine the central tendency of the data. When mapping the five levels of complexity in the questionnaire to the interval of complexity from 0 to 1, level 1 corresponds to a complexity between 0 and 0.2, level 2 corresponds to complexity interval (0.2, 0.4),…, level 5 corresponds to (0.8, 1.0). A visualization of this mapping is shown in Figure 4.7. The area which shows the consistency of the two methods is marked with color blue. As can be seen that, the majority of the points lie in this area. The outliers account for 15% of the total (3 of 20). Figure 4.8 compares the two methods by using mean ranks and the calculated complexity in Table 3.6. As can be seen that, scenarios with higher complexities usually have higher mean ranks, which means that they are evaluated by more experts with higher level of complexity. A linear relationship can be found between the results of these two methods. However, same as Figure 4.7 outliers exist as well. There is one scenario with a complexity larger than 0.9, but its mean rank is only the fifth highest. On the contrary, there is also a scenario, which according to the result of the metric is not very complex (complexity between 0.3 and 0.4), however is considered as the third most complex scenario by the experts. Ideally, the results of the metric will be more convincing if more points lie closely to the fitting line. In Figure 4.7, more dots should lie in the blue marked area. The goodness of this fit can be measured by norm of residuals, which is obtained with the following equation:

$$\text{normr} = \sqrt{r_1^2 + r_2^2 + \cdots + r_{20}^2} \tag{4.6}$$

r_i $(i = 1,2,\dots,20)$ is the difference between the observed value and fitted value of the i-th scenario. For the calculation of complexity with equal weighting factors, the norm of residuals is 0.4818.

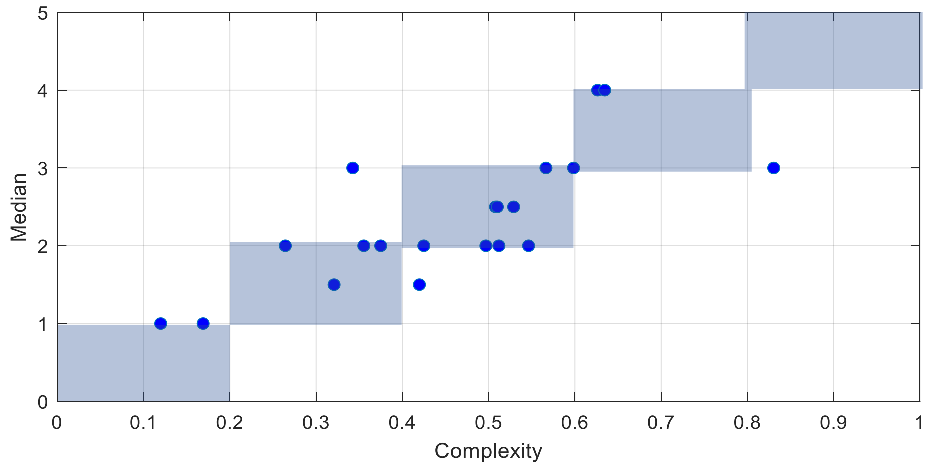


Figure 4.7: Comparison of results evaluated by experts and by metric

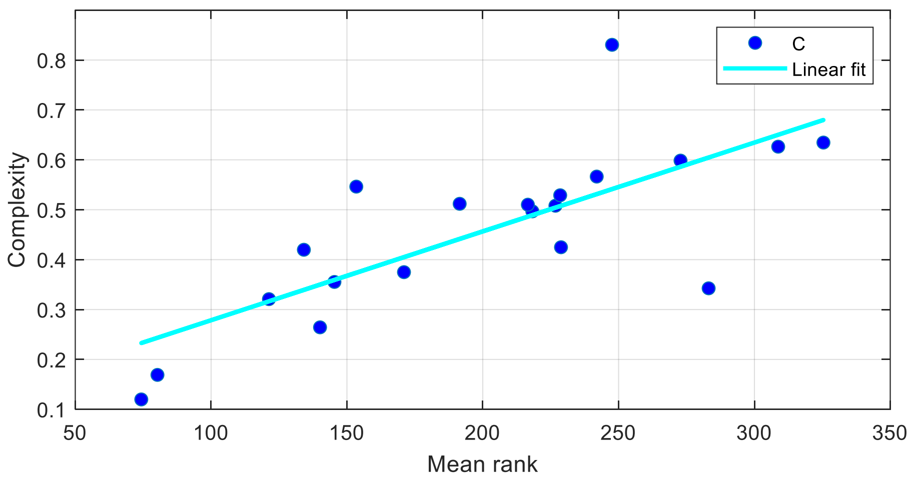


Figure 4.8: Comparison between the results of questionnaire and of metric

4.2 Determination of weighting factors

As already discussed in section 3.3.1. Some factors have strong influence on complexity, while some have very weak influence. The purpose of adjusting the weighting factors is to make the factors, which are very important according to the opinions of experts, very influential on the complexity, and the factors which are not important should have small impact on complexity.

According to the results of the second and third part of the questionnaire, only the difference between factor 6 and 1 and factor 4 and 1 are considered significant, namely for the majority of influence factors, their difference with respect to importance degree for complexity is not very large. Therefore, equal weighting factors are selected as the initial condition for the variation of weighting factors. In total 121 combinations of weighting factors are created for the evaluation of complexity. The sum of weighting factors of each group is equal to one. The norm of residuals of each combination is shown in Figure 4.9. The 1st point is the initial state (equal weighting factors). These 121 combinations are divided into 8 groups marked in color yellow in the graph. The detailed information of each group can also be seen in Table 4.9. The first column of the table is the group number. The second column shows the corresponding range of combination of each group. Under the third column it is described how weighting factors are distributed among factors within each group.

In the 1st group, the weighting factor of influence factor 1 w_1 varies from equal factor 0.0769 (equal weighting factor) slowly to 0, while the rest of the influence factors share the same weighting factors w_r. This corresponds to the first 8 combinations of weighting factors. As can be seen from the graph that, as w_1 gets smaller, the norm of residuals increases instead of decreasing. The same situation happens in the 2nd, 3rd and 4th group for influence factor 12, 9 and 7, which are the ones with lowest mean ranks (Figure 4.2). Similar actions can be done for the four factors with highest mean ranks (factor 6, 4, 5 and 11), the weighting factor of this factor varies from 0.0769 to 0.3, the rest of the influence factors are equally weighted (Group 5 to 8 in Figure 4.9). It is noticeable that when increasing weighting factor of factor 5, norm of residuals increases dramatically. One explanation for this is that, according to the result in 3.3.1 factor 6 is the most influential factor. When same degree of variance of an input happens, the variance at output is supposed to be the largest if this variance is caused by the most influential factor, in this case caused by factor 6.

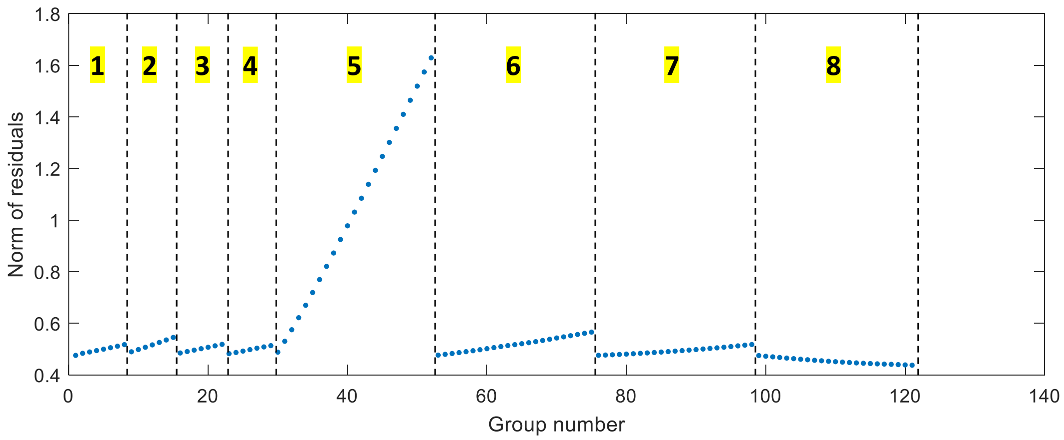


Figure 4.9: Norm of residuals of different combinations of weighting factor

The purpose is that, the most important factors should also be the most influential ones. For factor 6 it is already the case. It has the highest mean rank and is considered by the experts the most important factor. At the same time, it is the most influential one as well. Therefore, the weighting factor of factor 6 can remain unchanged. When comparing Figure 3.11 and Figure 4.5, it is noticeable that, factor 4 is supposed to be the second most important factor, however according to the results of sensitivity analysis, its influence on complexity is only in a medium level. Factor 1 and Factor 12 are the two least important factors, but their influence on the complexity is larger than factor 4. Therefore, in the first step, the weighting factor of factor 4 will be increased and factor 1 and 12 will be less weighted, so that factor 4 can have large influence on the result and factor 1 and 12 have less influence. The applied weighting factor in the first try is [0.01, 0.087, 0.087, 0.1, 0.087, 0.077, 0.087, 0.087, 0.087, 0.087, 0.1, 0.02, 0,084]. With this combination of weighting factors, the sensitivity analysis is conducted for track 7 and 8 (Since the results of sensitivity analysis of different tracks are highly consistent (Figure 3.11), the analysis will be conducted here for only two tracks instead of four to save computing time). The results are shown in Figure 4.10. As can be seen that, the rank of factor 4 has been improved, and the influence level of factor 1 and 12 is reduced. When using this combination of weighting factors to calculate the complexity of selected scenarios in the questionnaire and comparing the result with the evaluation of experts, the norm of residuals of the fitting line is increased to 0.5367 (Figure 4.11). The reason for increase of norm of residuals is that, when observing Figure 4.9, the decrease of weighting factor of factor 1 and 12, the increase of weighting factor of factor 4 all leads to increase of norm of residuals, only increase of weighting factor of factor 11 can reduce this level. The weighting factors of the rest factors remain almost the same, therefore the cause of increase of norm of residuals is more than that of decrease.

Table 4.9: Different combinations of weighting factors

Nr	Combi. range	Description
1	1 - 8	$w_1 \in \{0.0769, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0\}, w_r = (1 - w_1)/12$
2	9 - 15	$w_{12} \in \{0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0\}, w_r = (1 - w_{12})/12$
3	16 - 22	$w_9 \in \{0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0\}, w_r = (1 - w_9)/12$
4	23 - 29	$w_7 \in \{0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0\}, w_r = (1 - w_7)/12$
5	30 - 52	$w_6 \in \{0.08, 0.09, 0.10, \ldots, 0.28, 0.29, 3.0\}, w_r = (1 - w_6)/12$
6	53 - 75	$w_4 \in \{0.08, 0.09, 0.10, \ldots, 0.28, 0.29, 3.0\}, w_r = (1 - w_4)/12$
7	76 - 98	$w_5 \in \{0.08, 0.09, 0.10, \ldots, 0.28, 0.29, 3.0\}, w_r = (1 - w_5)/12$
8	99 - 121	$w_{11} \in \{0.08, 0.09, 0.10, \ldots, 0.28, 0.29, 3.0\}, w_r = (1 - w_{11})/12$

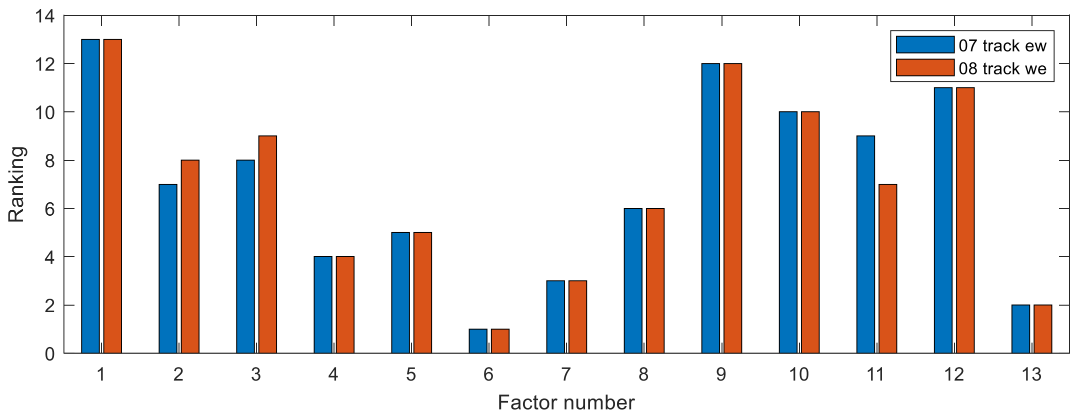


Figure 4.10: Sensitivity analysis with updated weighting factor (1)

The result in Figure 4.10 can then be compared with Figure 4.2 again, the weighting factors of factor 7 and 13 can be decreased, since they are not very important but have high level of influence. The weighting factors of factor 4, 5 and 11 can furthermore be increased to increase their influence level. The updated weighting factor is now: [0.01, 0.0667, 0.0667, 0.15, 0.1, 0.077 ,0.0667, 0.0667, 0.08, 0.0667, 0.2, 0.02, 0.03]. The result of sensitivity analysis is shown in Figure 4.12. When applying this combination of weighting factors to the calculation of complexity of selected scenarios in the questionnaire, the norm of residuals will be increased to

0.5906 after comparing with the evaluation of experts (Figure 4.13). As can be seen in Figure 4.12, the influence level of factor 4, 5 and 11 is improved further. The influence level of 7 and 13 is decreased.

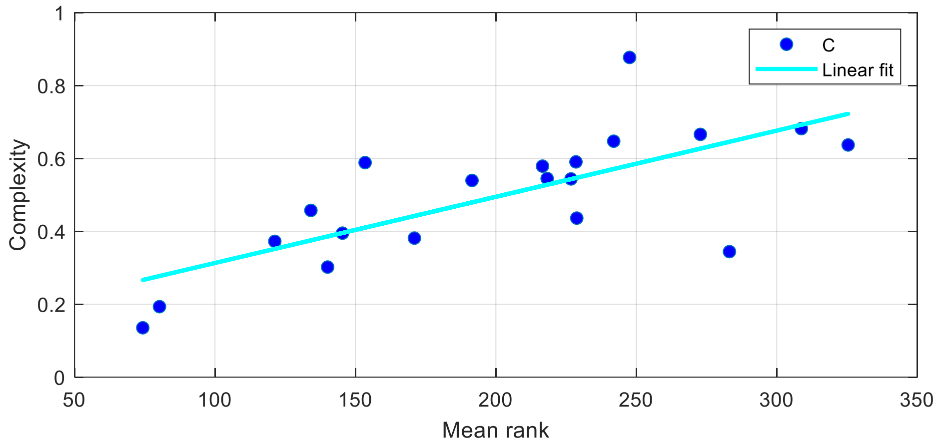


Figure 4.11: Comparison of evaluations by experts and by metric using updated weighting factor (1)

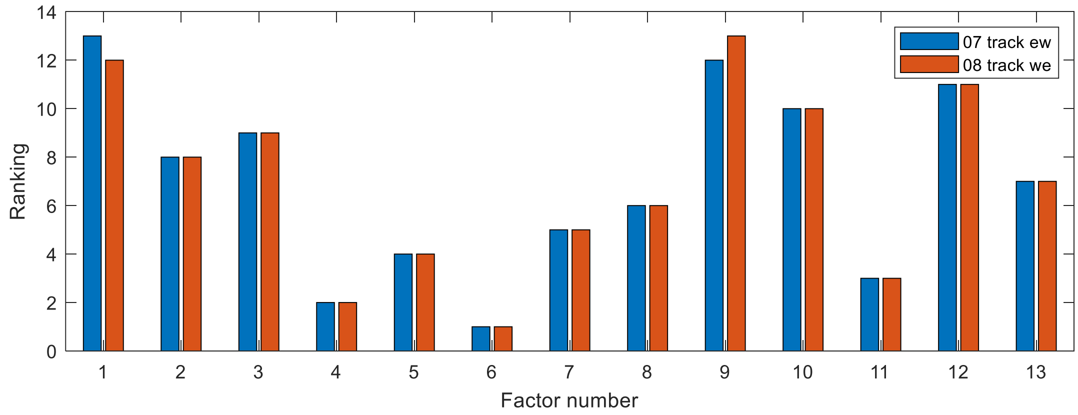


Figure 4.12: Sensitivity analysis with updated weighting factor (2)

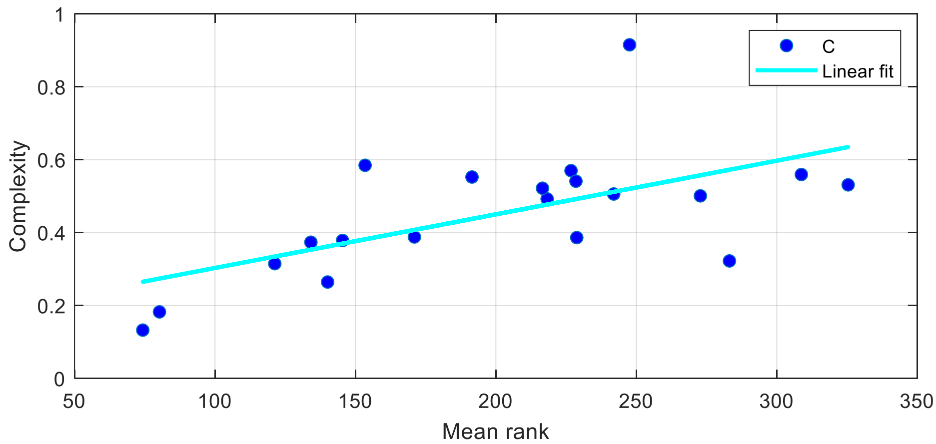


Figure 4.13: Comparison of evaluations by experts and by metric using updated weighting factor (2)

Further adjustment of weighting factors is possible, so that the ranking of influence level is coincident with the ranking of importance degree. The following weighting factors are adapted after making finer adjustments: [0.01, 0.08, 0.08, 0.15, 0.12, 0.08, 0.05, 0.05, 0.08, 0.08, 0.17, 0.02, 0.03]. The result of sensitivity analysis is displayed in Figure 4.14. As can be seen that, except factor 7, the ranking of influence level of other factors is almost coincident with the ranking of their importance degree. However, the norm of residuals of both evaluations is further increased to 0.6120 (Figure 4.15).

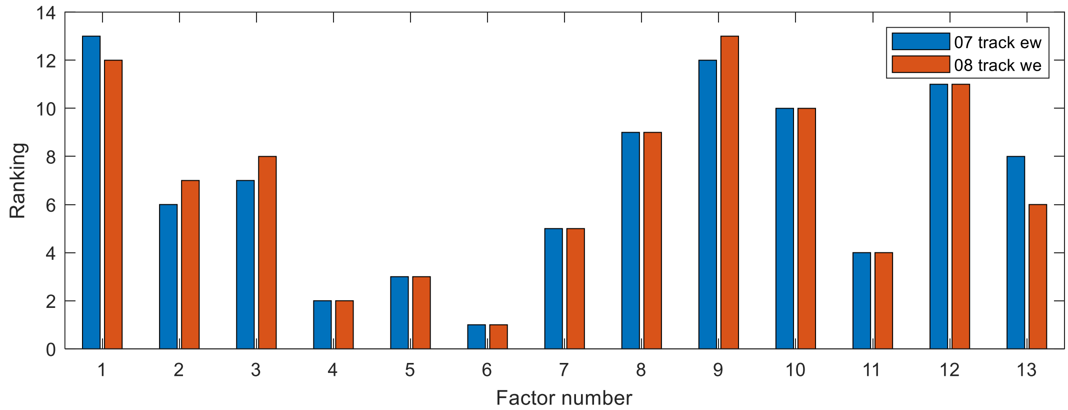


Figure 4.14: Sensitivity analysis with updated weighting factor (3)

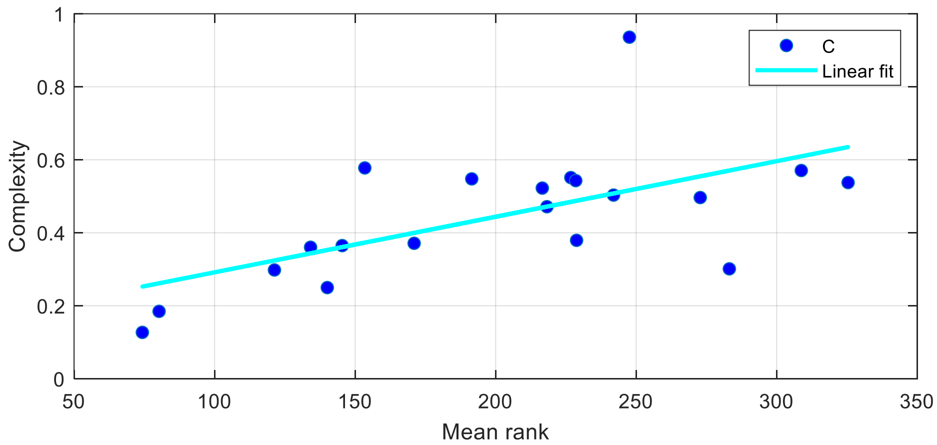


Figure 4.15: Comparison of evaluations by experts and by metric using updated weighting factor (3)

The question, which combination of weighting factors should be used, depends on the emphasized criterium. If the evaluation of experts is reliable and the results of the metric should be consistent with the opinions of experts, then the first combination of weighting factors can be adopted. With this combination the influence level of some important factors is improved and the difference between results of two evaluations will not be very large, which is a good compromise of both sides.

5 Summary

This work aims to perfect the metric for quantitative evaluation of traffic situation complexity from the perspective of automated ego-vehicle.

It has improved some shortcomings existing in the previous work. For the influence factor “connectivity” the connections between ego-vehicles and surrounding vehicles in diagonal directions are taken into consideration as well. The normalization of values of some factors is corrected. For factor “number of possible actions of ego-vehicle”, a piecewise function is used for normalization. The factor has the highest normalized value when ego-vehicle has 2 or 3 possible actions instead of just 1, which is also the opinions of the experts according to their comments in the questioonnaire, that the situation is more complex when the number of possible actions is somewhere in between. Because in this case more efforts is required for the decision making process. For the normalizeation of factor “time-gap”, the maximal value of time-gap between ego-vehicle and a surrounding vehicle (with one of the following five labels: “leftfollow”, “rightfollow”, “leftpre”, “middlepre”, “rightpre”) is used instead of the average value. Since very rare that more that three of the five sectors are occupied, the normalized value is usually very small when using the average value.

This work has added three more factors to the original ten for the evaluation of complexity. One is a scene-based factor “time-to-brake”, which is the reamining time for ego-vehicle to take an emergency brake if the preceding vehicle (vehicle labeled “middlepre”) comes to a sudden halt. This factor is included so that the criticality of a traffic situation can be reflected. The other two factors are scenario-based, which means instead of being calculated in each scene, they are calculated for each scenario and the values remain the same throughout the scenario. These factors are “number of actions of ego-vehicle/surrounding vehicle”. These two factors can indicate the dynamic of the traffic flow. If large number of actions is involved, this means that the vehicle in a situation where it has to frequently accelerate and decelerate and is probably in a traffic jam.

In the previous work the complexity is calculated with help of a linear model and equal weighting factors of all influence factors without any argumentation. This works studies the relationship between different influence factors and attempts to offer argumentation for the determination of weighting factors. A sensitivity anaylsis is conducted first with equal weighting factors to see how influential each influence factor is on the complexity. Sigma-normalized derivative is used as a measurement. Factor “types of surrounding vehicles”, “connectivity”, ”deviation” and “possible number of actions of surrounding vehicles” are the most influential ones. The concept “influence level” of a factor differs from its “importance degree” with respect to the assessment of complexity. The former depends on the distribution of the normalized values of respective factors and requires the knowledge and experience of experts. The importance degree of influence factors is determined based on the results of a questionnaire answered by experts. Based on the analysis of the results, factor 6 and factor 4 are considered as the most two important factors, factor 1 is then the least important. The difference between factor 6 or 4 and factor 1 is considered to be significant with respect to importance degree for the evaluation of complexity. The difference among other influence factors is not so significant. The final determination of weighting factors has taken the results of both sensitivity analysis and the questionnaire into consideration. Ideal situation or the purpose is that, the most influentail factors are the most important ones as well, at the same time, the scenario complexity evaluated by experts can be as close to the results of complexity evaluted by the develped metric as well. Sometimes these two conditions cannot be both satisfied and a compromise is necessary.

6 Discussion and outlook

The complexity in this work depends on the values of 13 influence factors, which are evaluated without consideration of state of the art of automated vehicle. For instance, the factor “types of surrounding vehicles” is included since different types of traffic participants have different characteristics of driving. However, according to the feedback of some experts in the questinnaire, automated vehicle is able to identify the type of a vehicle and therefore better predict its behavior. Or for factor “number of surrounding vehicles”, as the number of surrounding vehicles gets larger the computation time rises as well. However, for the automated vehicle which plans the trajectory based on the empty space around it the computation cost will not be high, if there are many surrounding vehicles. Similarly for factor “deviation of surrounding vehicles from the predicted trajectories”, the quality of prediction depends on the algorithm applied by the automated vehicle.

For determination of influence factors, it is a bit difficult to adapt the weighting factors by taking both influence level and importance degree of influence factors into consideration. Maybe except a linear model a more sophisticated model can be used. In this work it is preferred, that the results of the metric should be close to the ones of experts. This has the precondition, that the evaluation of experts is reliable. However, in the reality, it is possible that the evaluation of a person is very intuiitve and lacks of quantitative analyzation. Another problem is that, although the scenarios for the evaluation by experts are selected as representative as possible, but its total number is only 20. To obtain a more convincing result a larger number of scenarios are necessary for the validation.

List of figures

List of tables

References

[1] X. Yu, “Method for quantitative evaluation of traffic complexity on the highway,” Semester Thesis, Garching, Munich, Germany, 2019.

[2] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, “Definition der Begriffe Szene, Situation und Szenario für das automatisierte Fahren,” *Workshop Fahrerassistenz-systeme*, pp. 105–117, 2015.

[3] G. Bagschik, T. Menzel, and M. Maurer, Eds., *Ontology based Scene Creation for the Development of Automated Vehicles*: IEEE Intelligent Vehicles Symposium (IV), Suzhou, China, 2018.

[4] *HighD datatset.* [Online]. Available: https://www.highd-dataset.com/

[5] A. Saltelli *et al., Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley & Sons, Ltd, 2008.

[6] S. Noh and K. An, “Decision-Making Framework for Automated Driving in Highway Environments,” *IEEE Trans. Intell. Transport. Syst.*, vol. 19, no. 1, pp. 58–71, 2018, doi: 10.1109/TITS.2017.2691346.

[7] P. M. Junietz, “Microscopic and Macroscopic Risk Metrics for the Safety Validation of Automated Driving,” Dissertation, Darmstadt, Technische Universität, 2019.

[8] P. Schöner, L. Töttel, J. Doll, and M. J. Zöllner, Eds., *Predictive Trajectory Planning in Situations with Hidden Road Users Using Partially Observable Markov Decision Processes*: IEEE Intelligent Vehicles Symposium (IV), Paris, France, 2019.

[9] *Limesurvey.* [Online]. Available: https://www.limesurvey.org/de/

[10] P. Sedlmeier and F. Renkewitz, *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler*. Hallbergmons, Germany: Pearson, 2018.

[11] *5-point choice array type question.* [Online]. Available: https://manual.limesurvey.org/File:A-5_point.jpg

[12] S. S. Stevens, “On the theory of scales of measurement,” *Science*, 103 (2684), pp. 667–680, 1946.

[13] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, pp. 297–334, 1951.

[14] L. Vale, J. Silcock, and J. Rawles, “An economic evaluation of thrombolysis in a remote rural community,” *BMJ Clinical Research*, 314(7080), pp. 570–572, 1997.

[15] N. T. Richardson, “Konzeption und Langzeittest der Mensch-Maschine-Schnittstelle für das hochautomatisierte Fahren im Lkw,” Dissertation, München, Technische Universität, 2019.

[16] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis,” *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, Dec. 1952.

[17] R. A. Fisher, *Statistical Methods for research workers*. Edinburgh, UK: Oliver and Boyd, 1925.

[18] *Chi-squre distribution table.* [Online]. Available: https://people.smp.uq.edu.au/YoniNazarathy/stat_models_B_course_spring_07/distributions/chisqtab.pdf

[19] T. Pohlert, *PMCMR: Calculate Pairwise Multiple Comparisons of Mean Rank Sums (Version 4.0).* [Online]. Available: https://www.researchgate.net/publication/286242772_PMCMR_Calculate_Pairwise_Multiple_Comparisons_of_Mean_Rank_Sums_Version_40, 2015

[20] A. Dinno, “Nonparametric pairwise multiple comparisons in independent groups using Dunn's test,” *The Stata Journal*, no. 1, pp. 292–300, 2015.

[21] G. Cardillo, *Dunn's Test: a procedure for multiple, not parametric, comparison.* [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/12827

Appendix

Anhang A Appendix A List of scenarios

1: https://www.youtube.com/watch?v=9ojVyRYyiPk

2: https://www.youtube.com/watch?v=w4S5mX6DXis

3: https://www.youtube.com/watch?v=xnmLJZqJTag

4: https://www.youtube.com/watch?v=f-d44ulvFMo

5: https://www.youtube.com/watch?v=K2engJ8GETs

6: https://www.youtube.com/watch?v=3uV1l8jdsXY

7: https://www.youtube.com/watch?v=TQXLO8SpihE

8: https://www.youtube.com/watch?v=ibNUY904Cdk

9: https://www.youtube.com/watch?v=zm9Uj428oeU

10: https://www.youtube.com/watch?v=xGOdYsfN5hA

11: https://www.youtube.com/watch?v=ZnhlrH6uVHk

12: https://www.youtube.com/watch?v=WkWs-1os60s

13: https://www.youtube.com/watch?v=2V8Wb8wAzp0

14: https://www.youtube.com/watch?v=XfLf5S1gr_Q

15: https://www.youtube.com/watch?v=8ifA54kV7Bo

16: https://www.youtube.com/watch?v=Y1GkeFIcACk

17: https://www.youtube.com/watch?v=k26SAxLzXAM

18: https://www.youtube.com/watch?v=GDx7PADYuhQ

19: https://www.youtube.com/watch?v=3zDyRoRxtMk

20: https://www.youtube.com/watch?v=UjIoIs7uH8M