# **Vorangegangene Arbeiten:**

## Recep Furgan Torlak: *„Detektion der Bewegung von Verkehrsteilnehmern aus Positionsdaten“*

Eckpunkte der Arbeit:

* Datengewinnung (sequentielle GPS-Daten) und Merkmalsableitung (Geschwindigkeit, Beschleunigung, Orientierung, etc.)
* Datenabfrage aus der Simulation CARLA (mittels Python-Skript)
* Evaluierung echter GPS-Daten

Erkenntnisse:

* Aus reinen GPS-Daten lässt sich eine Vielzahl von Merkmalen ableiten, welche sich gut für die Klassifikation von Verkehrsteilnehmern eignen.
* Reale GPS-Daten können stark verrauscht/ungenau sein.

Probleme:

* Fehlender Realismus der Simulation (Sehr genaue Positionsdaten, kaum Unterschiede zwischen Fahrzeugen, …)
* Sehr starke Ungenauigkeiten der realen Daten

## Maximilian Sohl: *„Klassifizierung der Bewegungsmuster von Mobilfunkteilnehmern zur erweiterten Umfeldwahrnehmung autonomer Fahrzeuge“*

Eckpunkte der Arbeit:

* Klassifizierung von Verkehrsteilnehmern anhand gewonnenen Simulationsdaten (basierend auf der Vorarbeit Torlaks)
* Nutzung und Evaluierung der Klassifikationsverfahren SVM, Decision-Tree und K-Nearest-Neighbours

Erkenntnisse:

* Genutzte Klassifikationsverfahren (insbesondere SVM) zeigen eine gute Performanz mit unverfälschten (damit aber auch unrealistischen) Daten.

Probleme:

* Schlechte Performanz der Klassifikatoren mit verrauschten (realistischeren) Daten
* Beschränkung auf die drei oben genannten Verfahren
* Keine Datenvorverarbeitung/Merkmalsextraktion wie bspw. bei Fischer (dadurch sind die Daten sehr hochdimensional 🡪 Fluch der Dimensionalität)
* Lediglich Evaluierung mit Genauigkeit, kein anderes Qualitätsmaß (Precision, Recall, F1)
* Testdaten nach Maps getrennt 🡪 Gemeinsame Nutzung hätte zu mehr und diverseren Trainingsdaten geführt.

## Patrick Opitz: „Klassifizierung der Relevanz von Verkehrsteilnehmern zur Erweiterten Umfeldwahrnehmung Autonomer Fahrzeuge“

Eckpunkte der Arbeit:

* Bestimmung und Berechnung von Relevanzparametern (Abstand, TTC, THW)
* Entwicklung fester Szenarien + Implementierung dieser Szenarien

Erkenntnisse:

* Unklar…

Probleme:

* Ergebnisse der Arbiet machen nicht den Eindruck, dass diese gut (auf allgemeine Szenarien erweiterbar) implementiert sind. 🡪 keine Eignung für Integration in Pipeline/Datenintegration
* Ergebnisse und Vorgehen der Arbeit nicht wirklich fassbar.

## Dietmar Fischer: „Verwendung von Positionsdaten zur automatisierten Klassifizierung von Verkehrsteilnehmern mittels maschinellen Lernverfahren“

Eckpunkte der Arbeit:

* Beschäftigung mit weiteren Möglichkeiten zur Vorverarbeitung von Simulations- und Realdaten zur besseren Klassifizierung von Verkehrsteilnehmern (und deren Relevanz) mithilfe maschineller Lernverfahren
* Verkehrsteilnehmerklassifikation mittels Deep-Learning (Feed-Forward-Netz) und Q-Deep-Learning (Reinforcement-Learning)
* Relevanzklassifikation:
  + Bildung und Analyse von Risikoklassen mittels K-Means
  + Entwicklung eines FFN zur Vorhersage zukünftiger Geschwindigkeiten
  + Klassifikation der Relevanz auf Basis des Abstands und der zukünftigen relativen Geschwindigkeit mittels FF- und QDL-Netzes
* Umfangreiche, aber generische Implementierung, die sich gut in eine Pipeline/Dateninfrastruktur integrieren lassen dürfte

Erkenntnisse:

* Normale Feed-Forward-Netze lieferten sehr gute Ergebnisse bei der Klassifikation von Verkehrsteilnehmern innerhalb eines festen Kontexts (z.B. Training auf Simulationsdaten und Evaluierung auf Simulationsdaten)
* Das Reinforcement-Learning erwies sich als robust bei Änderung des Kontexts (Training auf Simulationsdaten und Evaluierung auf Realdaten)
* Risikoklasssen-Klassifikation zeigte eine schlechte Performanz

Probleme:

* Datenbasis besteht im Wesentlichen aus unverfälschten und zu erheblichen Teilen unrealistischen Simulationsdaten
* Schlechte Performanz der Risikoklassen-Klassifikation
* Relevanzdefinition ist vermutlich viel zu stark vereinfacht (nur Geschwindigkeit und Abstand sind mit Sicherheit nicht hinreichend)
* Beschränkung auf neuronale Netze (ggf. würden andere ML-Verfahren z.B. bei der Relevanzklassifikation zu besseren Ergebnissen führen)

# **Themensammlung:**

# **Möglichkeit 1: Pipeline**

***Aufbau einer Dateninfrastruktur für die Integration von Verfahren des maschinellen Lernens zur erweiterten Umfeldwahrnehmung autonomer Fahrzeuge.***

**Aufhänger: Im Zuge der Bearbeitung vorangegangener Arbeiten dieses Themenbereiches wurden einige Modelle (im Kontext des maschinellen Lernens) entwickelt, welche nun sinnvoll in eine Dateninfrastruktur/Pipeline integriert werden sollen.**

**Datengewinnung:**

***Simulationsdaten* können mit Hilfe der Verkehrssimulation CARLA und einem Python-Skript von Torlak gewonnen werden.**

***Reale Daten* können über eine App gesammelt werden, welche speziell für diesen Zweck durch die Hochschule Coburg entwickelt wurde.**

**Bereits entwickelte Modelle:**

* ***Verkehrsteilnehmer-Klassifikation***

|  |  |  |
| --- | --- | --- |
| **Klassifikationsverfahren** | **Arbeit** | **Bewertung der Performanz** |
| **K-Nearest-Neighbours** | **Sohl** | **gut auf unverrauschten Simulationsdaten, schlecht auf realistischeren Daten** |
| **Decision Tree** | **Sohl** | **gut auf unverrauschten Simulationsdaten, schlecht auf realistischeren Daten** |
| **Support Vector Machine** | **Sohl** | **gut auf unverrauschten Simulationsdaten, schlecht auf realistischeren Daten** |
| **Deep Q-Learning Netz** | **Fischer** | **Gut auf unverrauschten Simulationsdaten, aber eher schlecht auf Realdaten (Ursache ggf. Datenmenge)** |
| **Feed Forward Netz** | **Fischer** | **Gut auf unverrauschten Simulationsdaten und realen Daten** |

* ***Risikoklassen-Klassifikation***

|  |  |  |
| --- | --- | --- |
| **Klassifikationsverfahren** | **Arbeit** | **Bewertung der Performanz** |
| **Deep Q-Learning Netz** | **Fischer** | **Sehr schlecht (nur 38% Präzision auf Simulationsdaten) bei Bestimmung vorher geclusterter „Risikoklassen“** |
| **Feed Forward Netz** | **Fischer** | **Sehr schlecht (46% Präzision auf Simulationsdaten) bei Bestimmung vorher geclusterter „Risikoklassen“** |

* ***Geschwindigkeitsprognose***

|  |  |  |
| --- | --- | --- |
| **Klassifikationsverfahren** | **Arbeit** | **Bewertung der Performanz** |
| **Feed Forward Netz** | **Fischer** | **Im Allgemeinen recht gut für die Regression der unmittelbar (0,5s) folgenden Geschwindigkeit auf Basis einer aktuell durchgeführten Handlung (Bremsen/Beschleunigen). Datenbasis: unverrauschte Simulationsdaten.** |

**Forschungsfragen:**

**Wie lassen sich verschiedene maschinelle Lernverfahren kombinieren/einsetzen, um auf Basis erfasster Positionsdaten anderer Verkehrsteilnehmer die Umfeldwahrnehmung autonomer Fahrzeuge zu erweitern?**

**Was soll/kann durch die Klassifikation von Verkehrsteilnehmern (Art, Relevanz) erzielt werden? 🡪 „Welchen Zweck verfolgt das Ganze? Macht das in der Praxis überhaupt Sinn?“**

**Offene Probleme:**

**Viele der Modelle zeigen aktuell noch eine unzureichende Performanz, um aus diesen eine Pipeline aufzubauen. Problematisch ist besonders die Gefahr, dass sich Fehler in aufeinander aufbauenden Modellen aufaddieren, so dass eine Pipeline (bspw. bestehend aus Geschwindigkeitsprognose und Risikoklassen-Klassifikation) nochmal deutlich schlechter ist als ihre einzelnen Bausteine.**

**Gegebenenfalls kann die Performanz der einfacheren Modelle (K-Nearest, SVM und Decision-Tree) durch eine bessere Datenvorverarbeitung/Merkmalsextraktion deutlich verbessert werden.**

**Relevanzklassifikation ist noch nicht ansatzweise reif, um in eine Dateninfrastruktur eingebettet zu werden.**

# Möglichkeit 2: PoC mit Realdaten/realistischeren Simulationsdaten

***Einsatz von Verfahren des maschinellen Lernens zur Klassifikation von Verkehrsteilnehmern auf Basis realistischer/realer Positionsdaten.***

**Aufhänger: Bisher lag der Fokus der vorangegangenen Arbeiten darin, die entwickelten Modelle mit Hilfe von Daten aus der Verkehrssimulation CARLA zu trainieren und entsprechend zu validieren. Diese Daten spiegeln jedoch kaum bzw. nur unzureichend die Realität wider. Aus diesem Grund sollten Modelle, welche sich als vielversprechend erwiesen haben, mit einer größeren Menge realer Daten optimiert, trainiert und validiert werden. Eine Möglichkeit zur Datengewinnung bietet bspw. die App zur Sammlung realer Daten.**

**Forschungsfrage:**

**Wie können reale Positionsdaten, die Ungenauigkeiten und Rauschen aufweisen, so vorverarbeitet werden, dass sie sich gut für den Einsatz maschineller Lernverfahren eignen?**

**Welche Klassifikationsverfahren sind für die Klassifizierung von Verkehrsteilnehmern auf Basis realer Positionsdaten geeignet?**

**(Viele Klassifikationsmöglichkeiten wurden noch nicht probiert, z.B. Ensemble-Verfahren.)**

**Offene Probleme:**

**Erste Versuche mit verrauschten Simulationsdaten und realen Daten zeigen zum großen Teil schlechte Performanz.**

**Methodik:**

**Datenvorverarbeitung und Einsatz von Map-Matching-Verfahren um Ungenauigkeiten zu reduzieren/auszugleichen. Denkbar wäre auch der Einsatz einer linearen Ausgleichsrechnung oder eines Linearen-Regressionsverfahrens um die Daten zu „glätten“.**

**Erzeugung großer und möglichst realistischer Simulationsdaten oder großangelegte Sammlung von realen Daten für das Training von Klassifikatoren.**

**Evaluierung weiterer maschineller Klassifikationsverfahren, bspw. von Ensemble-Methoden.**

# Möglichkeit 3: Relevanzklassifikation

*Klassifikation der Relevanz von Verkehrsteilnehmern auf Basis sequenziell gesammelter Positionsdaten mit Hilfe von Verfahren des maschinellen Lernens.*

**Aufhänger: Alle vorangegangenen Ansätze, Relevanz mit Hilfe von Positionsdaten zu bestimmen, können als „unzureichend“ eingestuft werden, um sinnvoll in eine Dateninfrastruktur zur erweiterten Umfeldwahrnehmung integriert zu werden. Da die Relevanzklassifikation innerhalb dieses Bereichs jedoch ein entscheidender Faktor ist, sollte das Thema erneut aufgegriffen und durch die Untersuchung neuer und verschiedener Ansätze bearbeitet werden.**

**Forschungsfragen:**

**Was bedeutet „Relevanz“ im Kontext der erweiterten Umfeldwahrnehmung von autonomen Fahrzeugen?**

**Wie lässt sich „Relevanz“ messen/beurteilen?**

**Welche Verfahren (des maschinellen Lernens) können eingesetzt werden, um aus simulierten oder realen Positionsdaten Aussagen über die „Relevanz“ anderer Verkehrsteilnehmer abzuleiten?**

**Methodik:**

**Markov-Modelle**

**Ggf. können aus Positionsdaten weitere Infos abgeleitet werden. Hilfreich hierbei wäre bspw. die Nutzung einer Road-API, wie derer von Google (kostenpflichtig;** <https://developers.google.com/maps/documentation/roads/overview?hl=de>**). Zum Beispiel könnte damit festgestellt werden, ob sich Fahrzeuge (wahrscheinlich) auf derselben Straße oder angrenzenden Straßensegmenten befinden.**

**Ggf. existieren auch offene APIs, die für diesen Zweck ausreichend sind (weitere Recherchen nötig).**

* Road-APIs machen natürlich nur für Realdaten-Sinn. Allerdings kann die Grundidee, ob sich Fahrzeuge auf den gleichen oder angrenzenden Straßen befinden ggf. auch auf Carla übertragen werden.