

# Fallstudien II

---

Laura Kampmann, Christian Peters, Alina Stammen

12. Dezember 2020

## 1. Task I - Vorhersage der Datenrate

Gradient Boosted Trees

Regression mit ARMA-Fehlern

Validierung

## 2. Task II: Handover Vorhersage und Link Lifetime

Lösungsansatz Task II

## **Task I - Vorhersage der Datenrate**

---

# Task I - Vorhersage der Datenübertragungsrate

- Ziel: Evaluation von neuen *anticipatory vehicular communication systems* durch möglichst realitätsnahe Simulationen [2]  
⇒ Ansatz: *Data-Driven Network Simulation*
- Durch Machine Learning Modelle sollen möglichst realistische Vorhersagen der Datenraten generiert werden
- Hoffnung: Bessere Aussagekraft der Simulationen durch Einsatz echten Datenmaterials

# **Task I - Vorhersage der Datenrate**

---

**Gradient Boosted Trees**

# Gradient Boosted Trees

- Kann man aus vielen "schwachen" Lernern einen starken Lerner konstruieren?
  - ⇒ Ja, Boosting ist eines der mächtigsten Konzepte des Machine Learning [1]
- Kombination von einfachen CART Bäumen zu einem starken Ensemble
  - ⇒ Ähnlich zu Random Forest
- Der Unterschied zum Random Forest liegt im Training!

# Training von Gradient Boosted Trees

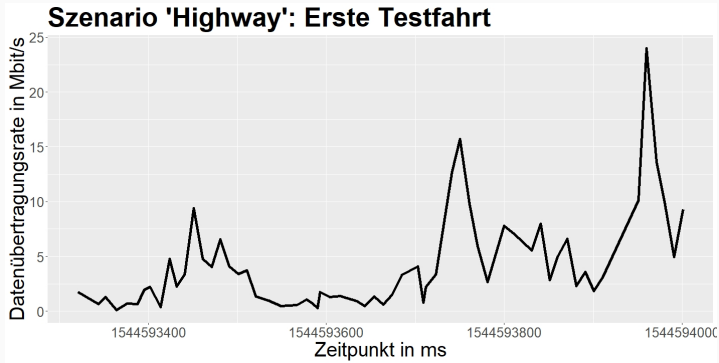
- Bäume werden nacheinander zum Ensemble hinzugefügt
- Jeder neue Baum versucht, die Schwächen seiner Vorgänger "auszubügeln"
  - ⇒ *Additives Training*
- Je mehr Bäume aufgenommen werden, desto geringer wird der Training-Error (das Modell wird aber komplexer)
  - ⇒ Kontrolle des *Bias-Variance Tradeoffs*
  - ⇒ Zusätzlich gibt es Regularisierungs-Parameter

# **Task I - Vorhersage der Datenrate**

---

**Regression mit ARMA-Fehlern**





**Figure 1:** Grafik der auf der ersten Testfahrt im Szenario “Highway” gemessenen Datenübertragungsrate.

- Zeitreihe  $y_1, \dots, y_n$  (Zielvariable)
- $k$  Zeitreihen  $x_{i,1}, \dots, x_{i,n}$  für  $i = 1, \dots, k$  (Einflussvariablen)

## Lineares Regressionsmodell

$y_t = c + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \epsilon_t$  mit Fehler  $\epsilon_t$  und Konstante  $c$

Annahmen an Fehler:

- $E((\epsilon_1, \dots, \epsilon_n)^T) = 0$
- $Cov((\epsilon_1, \dots, \epsilon_n)^T) = \sigma^2 \mathbb{1}_n$

**Annahmen sind in unserer Situation nicht einhaltbar!**

**ARMA( $p$ ,  $q$ )**: Zusammengesetztes Modell aus

- $AR(p)$  (Auto Regressive): Linearkombination aus
  - $p$  vorherige Beobachtungen,
  - Konstante
  - Fehler
- $MA(q)$  (Moving Average): Linearkombination aus
  - $q$  vorherige Fehler
  - Konstante
  - aktueller Fehler

## Anwendung auf Regressionsfehler

Erinnerung: Fehler  $(\epsilon_1, \dots, \epsilon_n)$  des linearen Modells sind autokorreliert  $\Rightarrow$  erfüllen Voraussetzungen nicht

Lösung: Wende ARMA-Modell auf Fehler an

**Modellgleichung Regression mit ARMA-Fehlern:**

$$y_t = c + \sum_{i=1}^k \beta_i x_{i,t} + \underbrace{\sum_{j=1}^p \phi_j \epsilon_{t-j}}_{\text{vergangene Fehler LM}} + \underbrace{\sum_{k=1}^q \theta_k \epsilon_{t-k}}_{\text{vergangene Fehler ARMA}} + \epsilon_t$$

# **Task I - Vorhersage der Datenrate**

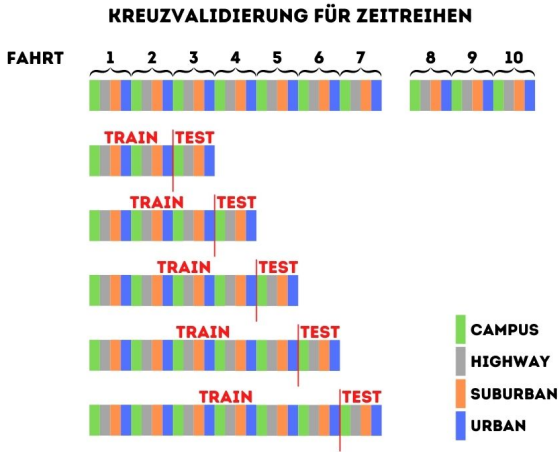
---

**Validierung**

## **k-fache Kreuzvalidierung**

- beachtet Abhängigkeit der Datenpunkte nicht
- zerstört zeitliche Komponente
- verwendet eventuell zukünftige Beobachtungen für Prognose der Gegenwart

⇒ **Kreuzvalidierung für Zeitreihen**



**Figure 2:** Einteilungen in Trainings- und Testdatensätze bei der Kreuzvalidierung für Zeitreihen.

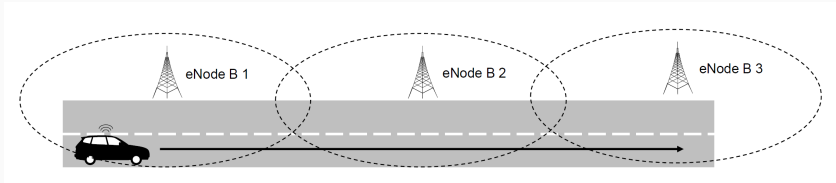
## **Task II: Handover Vorhersage und Link Lifetime**

---



## Vorhersage des Handovers und Link Lifetime

- Vergleich des RSRP Wertes zur verbundenen Zelle sowie zu den Nachbarzellen
- Vorhersage des Handovers durch Angabe der Link Lifetime



# **Task II: Handover Vorhersage und Link Lifetime**

---

## **Lösungsansatz Task II**

# Lösungsansatz Task II

**Idee:** Prädiktionsmodell für Link Lifetime mit Einfluss des RSRP der verbunden sowie der Nachbarzellen

→ Datentransformation nötig

- Anpassen der RSRP Messwerte in "Cells" an RSRP Werte in "Context"
- Cell Id → eNodeB
- eNodeB Wechsel → Response Variable Link Lifetime

time_s	rsrp_dbm	ci	scenario	provider	enodeb	drive_id	rsrp_neighbor	link_lifetime
0.06	-98	13828122	campus	o2	54016	1	-99	18.01
1.07	-101	13828122	campus	o2	54016	1	-104	17.00
2.07	-101	13828122	campus	o2	54016	1	-104	16.00
3.07	-94	13828122	campus	o2	54016	1	-100	15.00
4.07	-94	13828122	campus	o2	54016	1	-100	14.00

- Anwendung des Prädiktionsmodells XGBoost um Link Lifetime vorherzusagen
- Validierung analog zu Task I mit Zeitreihenkreuzvalidierung



T. Hastie, R. Tibshirani, and J. Friedman.

**The elements of statistical learning: data mining, inference and prediction.**

Springer, 2 edition, 2009.



B. Sliwa and C. Wietfeld.

**Data-driven network simulation for performance analysis of anticipatory vehicular communication systems.**

*IEEE Access*, 7:172638–172653, 2019.