

Prognose von Fahrgastwechselzeiten mit Live-Daten aus dem Betrieb

Die Weiterentwicklung von Fahrerassistenzsystemen bietet das Potenzial, den Bahnbetrieb zugunsten der Pünktlichkeit und des Energiebedarfes zu optimieren. Um frühzeitig Schwankungen im Betriebsgeschehen berücksichtigen zu können, gewinnt die Berechnung von Vorhersagen zunehmend an Bedeutung. In diesem Beitrag liegt der Fokus auf der Prognose von Fahrgastwechselzeiten basierend auf tagesaktuellen Zähldaten.



Motivation

Schwankungen im Fahrgastaufkommen und folglich veränderte Ein- und Ausstiegszeiten können zu Haltezeitabweichungen führen und sich somit auf die Pünktlichkeit und Kapazität im Bahnbetrieb auswirken (vgl. [1] und [2]). Mithilfe von Haltezeitprognosen können im Live-Betrieb optimierte Fahrtrajektorien berechnet und somit die Fahrweise in Hinblick auf Pünktlichkeit und

Energiebedarf entsprechend angepasst bzw. optimiert werden.

Im mFund-Projekt FASaN [3] wird, aufbauend auf vorherigen Arbeiten [4], ein vernetztes Fahrerassistenzsystem (vFAS) entwickelt und erprobt. Ein Teilziel des Projekts ist die Berechnung von Echtzeit-Prognosen für die Fahrgastwechselzeit. In Abhängigkeit von der Prognose werden anschließend optimierte Fahrtrajektorien für einzelne Streckenabschnitte berech-



Maike Krips

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrssystemtechnik, Wissenschaftliche Mitarbeiterin in der Gruppe Bahnbetrieb in der Abteilung Design & Bewertung von Mobilitätslösungen
maike.krips@dlr.de



Dr.-Ing. Christian Meirich

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrssystemtechnik, Gruppenleiter Bahnbetrieb in der Abteilung Design & Bewertung von Mobilitätslösungen
christian.meirich@dlr.de



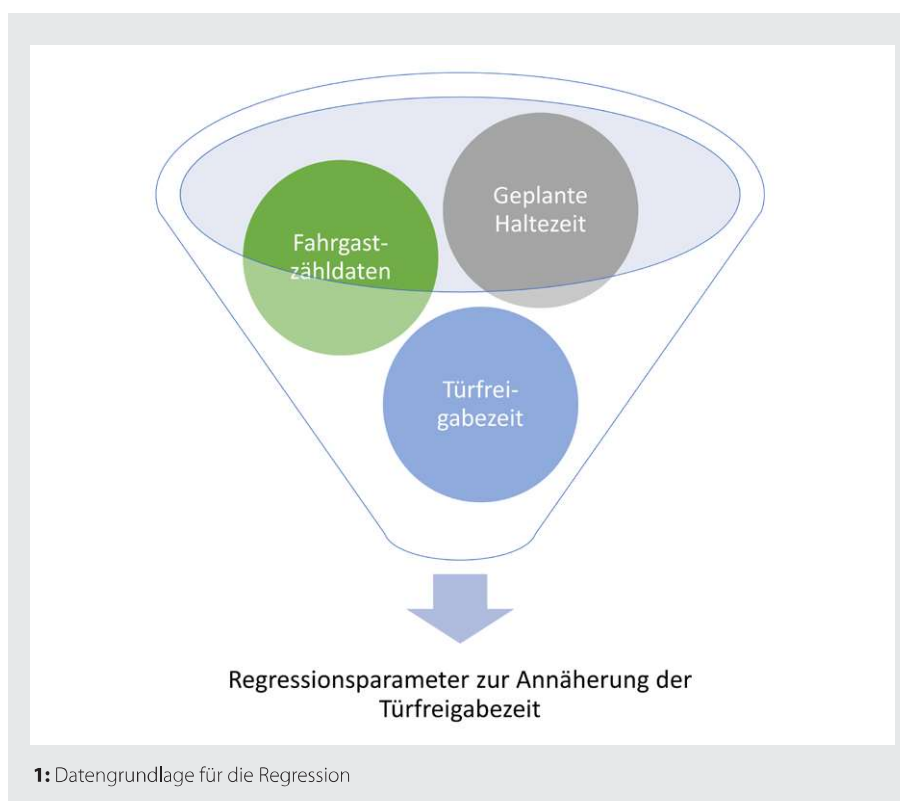
Dipl.-Ing. Patrick Reiher

INAVET – Institut für angewandte Verkehrstelematik GmbH, Projektleiter Vollbahn
patrick.reiher@inavet.de



Dipl.-Ing. Felix Zöllner

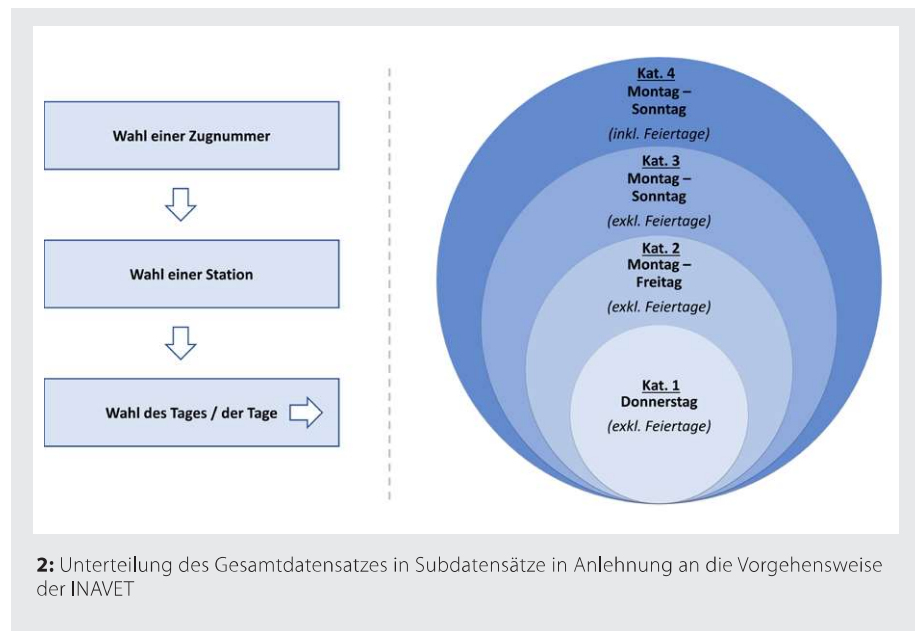
INAVET – Institut für angewandte Verkehrstelematik GmbH, Projektgenieur
felix.zoellner@inavet.de



net und mithilfe von Fahrempfehlungen (FE) an die Triebfahrzeugführenden (Tf) übermittelt. Ist die prognostizierte Fahrgastwechselzeit länger als die im Fahrplan vorgesehene Haltezeit, so ist mit dem Aufbau einer Verspätung zu rechnen. In diesem Fall führt das Fahren mit maximal zulässiger Geschwindigkeit dazu, dass der Halt möglichst früh erreicht wird und bestenfalls mehr Zeit für den Ein- und Ausstieg zur Verfügung steht. Dies wirkt sich wiederum positiv auf den Abbau einer möglichen Verspätung aus, führt jedoch zu einem erhöhten Energiebedarf. Liegt die Dauer der prognostizierten Fahrgastwechselzeit hingegen unter der ursprünglich eingeplanten Haltezeit, so ist kein Verspätungsaufbau zu erwarten. In diesem Fall ist unter Einbezug der Fahrplanpufferzeiten eine spätere Ankunft ausreichend, um die Pünktlichkeit aufrecht zu erhalten. Es besteht daher die Möglichkeit, vor Erreichen des nächsten Halts die FE „frühzeitiges Ausrollen“ durch das vFAS auszugeben. Dadurch kann die Geschwindigkeit des Zuges verringert und Energie eingespart werden.

Zur Prognose von Haltezeiten existieren verschiedene Ansätze, die sich hauptsächlich in der Datengrundlage sowie im verwendeten statistischen Modell unterscheiden. In Bezug auf die Daten können sowohl historisch gemessene als auch tagesaktuelle Werte verwendet werden. Meist wird ein historischer Datensatz genutzt, um wiederkehrende Zusammenhänge zwischen den Eingangsparametern und der gemessenen Haltezeit zu identifizieren und diese z.B. in Form von Regressionskurven abzubilden. Anschließend werden, ausgehend vom berechneten Modell, für tagesaktuelle Daten individuelle Echtzeit-Prognosen berechnet (vgl. [5], [6]).

In [6] wurden als Datengrundlage die Belegungen von Streckenblöcken inklusive der Länge, Position und Verspätung einzelner Züge in den Niederlanden genutzt. Um zeitliche Schwankungen im Tagesverlauf zu berücksichtigen, wurden die Daten in die beiden Kategorien „Spitzenzeit“ und „keine Spitzenzeit“ unterteilt. Dabei konnte gezeigt werden, dass die Haltezeiten innerhalb der zweiten Kategorie signifikant kürzer waren. Beim Vergleich verschiedener Modelle wurden dementsprechend nicht der (globale) Gesamtdatensatz, sondern stattdessen (lokale) Subdatensätze zur Berechnung von Regressionsmodellen verwendet. In dem vorgestellten Anwendungsfall haben sich zur Berechnung von Prognosen anhand des Gesamtdatensatzes



der Random-Forest-Algorithmus und in Bezug auf Subdatensätze ein lineares Regressionsmodell als am besten geeignet erwiesen. Bei der vorgestellten Methode wurden die Haltezeitdaten jedoch nicht sensorisch erfasst, sondern mithilfe von Blockbelegungszeiten im Bahnhof abgeschätzt, wodurch Abweichungen zur tatsächlichen Haltezeit entstehen können.

Der Einfluss von Fahrgastzahlen auf die Haltezeit wurde in [7] untersucht. Dabei konnte für Nahverkehrszüge in Stockholm gezeigt werden, dass mit steigender Fahrgastanzahl die Häufigkeit von Verspätungen zunimmt, wobei der Zusammenhang im betrachteten Untersuchungsraum nicht linear war. Die Verspätungsverteilung war zusätzlich von der Station und der Länge der dort eingeplanten Haltezeit abhängig.

Im Projekt FASaN besteht die Datengrundlage aus Informationen zur ge-

In Abhängigkeit von der Prognose werden [...] optimierte Fahrtrajektorien für einzelne Streckenabschnitte berechnet und mithilfe von Fahrempfehlungen an die Triebfahrzeugführenden übermittelt.

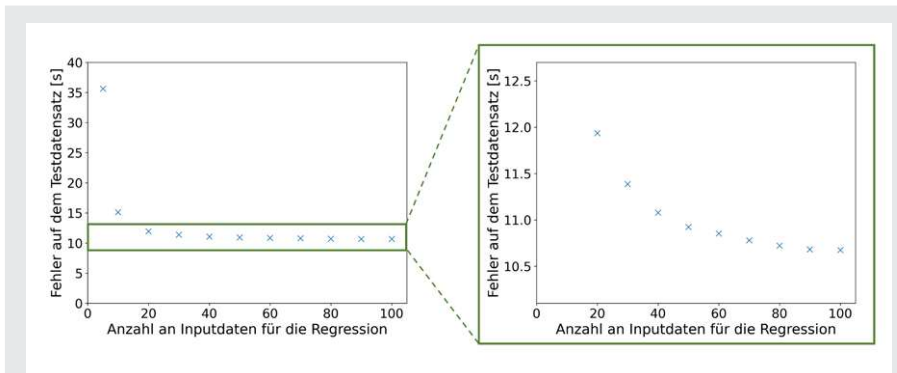
planten Haltezeit sowie zur gemessenen Türfreigabezeit, sprich der Zeit, in der die Türen des Zuges geöffnet sind. Zusätzlich werden diese Daten durch kameragestützte Fahrgastzählungen an den Türen erweitert, wodurch Informationen zur Anzahl an eingestiegenen, ausgestiegenen sowie sich im Zug befindlichen Passagieren zur Verfügung stehen (siehe Bild 1). Mit allen genannten Daten wird ein Regressionsmodell zur Schätzung der tatsächlichen Türfreigabezeit mittels Huber Regression aufgestellt. Im Projekt ist die Ostdeutsche Eisenbahn GmbH (ODEG) als Eisenbahnverkehrsunternehmen beteiligt. Die analysierten Daten stammen aus Fahrten auf dem ODEG-Liniennetz.

In diesem Artikel werden drei Fragestellungen näher betrachtet:

1. Wie viele Daten werden mindestens benötigt, um eine aussagekräftige Prognose berechnen zu können?
2. Lässt sich die Prognosegenauigkeit durch das Betrachten lokaler Subdatensätze verbessern?
3. Wie groß ist der Einfluss von Ungenauigkeiten bei der Prognose auf den Aufbau von Verspätungen und den Energiebedarf?

Methodik

Die im Projekt entwickelte Methodik basiert auf drei Schritten: Zunächst wird eine Datenvorverarbeitung durchgeführt, durch die Ausreißer, wie beispielsweise Messfeh-



3: Einfluss der Anzahl an Dateneinträgen auf das Regressionsergebnis bei 319 betrachteten Zugnummer-Station-Kombinationen

ler, aus dem Datensatz entfernt werden. So werden ausschließlich Züge mit definierten Mindest- und Maximalhaltezeitdauern berücksichtigt. Durch letztere werden zudem Züge mit stark verspäteter Abfahrt aus dem Datensatz entfernt. Bei ihnen ist die Fahrgastwechselzeit vermutlich weniger von der Fahrgastanzahl, sondern vielmehr von äußeren Einflussfaktoren wie beispielsweise dem Warten auf Anschlussreisende abhängig.

Im zweiten Schritt wird der Gesamtdatensatz in kleinere, lokale Subdatensätze unterteilt, die ähnliche Halte abbilden (siehe Bild 2, linke Seite). Dabei werden zunächst die einzelnen Zugnummern berücksichtigt, da sie sowohl uhrzeit- als auch richtungsspezifisch sind. Nachdem für jede Zugnummer ein Subdatensatz mit den zugehörigen Halten erstellt wurde, erfolgt anschließend eine weitere, feinere Unterteilung in Abhängigkeit der angefahrenen Station. Danach werden mögliche Kombinationen aus Tagen (z.B. Werktag/Wochenende) berücksichtigt. In Bild 2 sind auf der rechten Seite beispielhaft die zeitlichen Auswahlmöglichkeiten für eine Zugfahrt an einem Donnerstag gezeigt. Bei der Auswahl wird zunächst der Subdatensatz mit der höchsten Ähnlichkeit zum betrachte-

Bei einer geringen prognostizierten Fahrgastwechselzeit erhöht sich die verfügbare Fahrzeit und es wird ein frühes Ausrollen vom FAS empfohlen [...].

ten Tag priorisiert, spricht der Datensatz über alle Zugfahrten mit derselben Zugnummer und Station an ebenfalls einem Donnerstag (Kategorie 1). Ist dieser Datensatz zu klein, um eine robuste Regressionskurve berechnen zu können, so kann auf die nächsthöhere Kategorie 2 (Montag bis Freitag, exklusive Feiertage) zurückgegriffen werden.

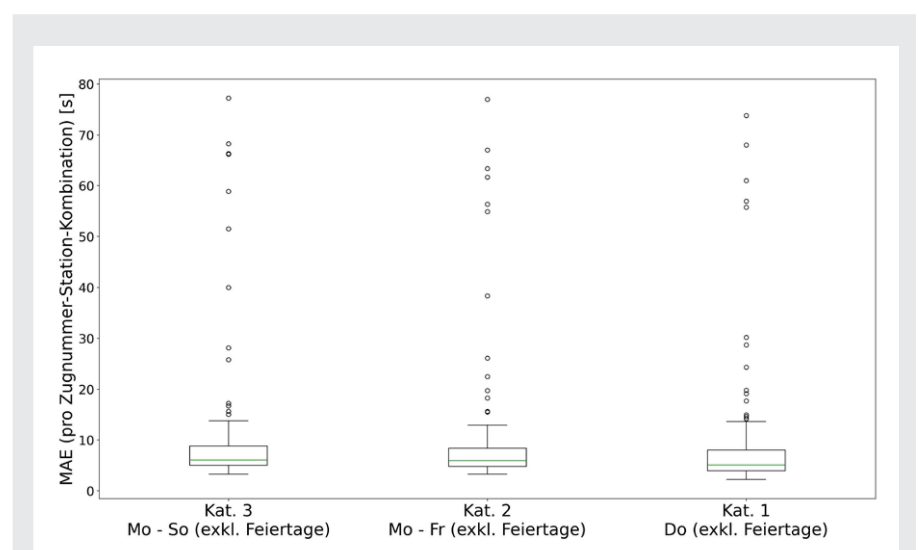
Im letzten Schritt der Methodik wird mithilfe der Huber Regression [8] eine Annäherung an die Fahrgastwechselzeit berechnet. In Vorarbeiten hat dieses Verfahren im Vergleich zu weiteren untersuchten Modellen (Lineare, RANSAC- und Decision Tree Regression) die geringsten Abweichungen zur betrachteten Zielgröße aufgewiesen. Bei der Huber Regression handelt es sich um eine Variante der linearen Regression,

die vergleichsweise robust gegenüber Ausreißern im Datensatz ist (vgl. [8]).

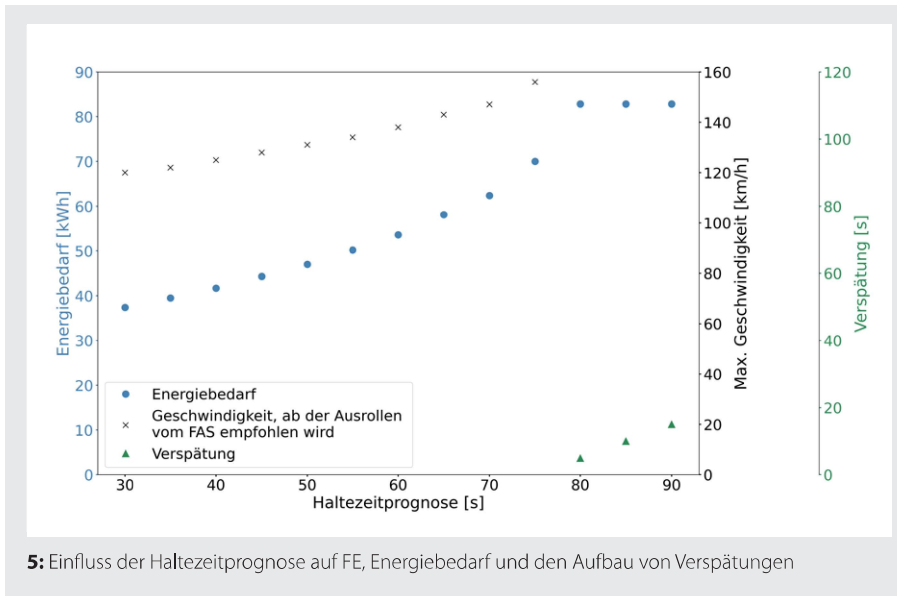
Zur Beantwortung der ersten Fragestellung werden für jede vorhandene Kombination aus Zugnummer und Station Subdatensätze gebildet. Alle Subdatensätze, die mindestens 150 Einträge enthalten, werden anschließend in einen Trainings- und einen Testdatensatz mit dem Verhältnis von 2/3 zu 1/3 unterteilt. Ersterer dient dazu, die Regressionsparameter zu berechnen, während mithilfe des letzteren die Abweichung zwischen prognostizierten und tatsächlichen gemessenen Daten analysiert wird. Um den Einfluss der Datenanzahl zu untersuchen, wird der Trainingsdatensatz dabei sukzessive vergrößert.

In Bezug auf die zweite Fragestellung zur Verwendung lokaler Datensätze werden für die in Bild 2 definierten Subdatensätze separate Regressionskurven berechnet und anschließend ebenfalls die Differenz zwischen angenäherten und gemessenen Werten ermittelt. Als Vergleichsgröße dient hierbei der mittlere absolute Fehler (englisch: Mean Absolute Error, kurz: MAE).

Der Einfluss von Prognoseabweichungen auf den Betrieb (Fragestellung 3) wird simulationsbasiert in Form einer Sensitivitätsanalyse untersucht. Dabei werden beispielhaft Fahrten zwischen zwei Haltepunkten auf einem ausgewählten Streckenabschnitt untersucht. Die gemäß Fahrplan zur Verfügung stehende Fahrzeit beträgt im betrachteten Fall 300 s. Für den zweiten Halt werden zeitlich variierende



4: Analyse des Einflusses von Subdatensätzen auf den mittleren absoluten Fehler bei 150 betrachteten Zugnummer-Station-Kombinationen



5: Einfluss der Haltezeitprognose auf FE, Energiebedarf und den Aufbau von Verspätungen

Prognosen für den Fahrgastwechsel angenommen, wodurch sich die jeweilige FE ändert: Bei einer geringen prognostizierten Fahrgastwechselzeit erhöht sich die verfügbare Fahrzeit und es wird ein frühes Ausrollen vom FAS empfohlen, wohingegen im umgekehrten Fall ein spätes oder ggf. kein Ausrollen empfohlen wird. Die zugehörigen Auswirkungen auf den Betrieb in Form von Verspätungen sowie die Energiebedarfe werden mithilfe des Simulationsmodells berechnet. Der angegebene Bedarf berechnet sich dabei aus der Energieaufnahme und Rekuperation.

Ergebnisse

Bei der Analyse der erforderlichen Anzahl an Dateneinträgen für die Prognoseberechnung wurden insgesamt 319 Kombinationen aus Zugnummer und Station berücksichtigt. In Bild 3 ist der MAE auf dem Testdatensatz über alle betrachteten Kombinationen gemittelt dargestellt. Mit steigender Anzahl an Inputdaten sinkt der durchschnittliche Fehler. Konkret verringert er sich zunächst von ca. 14,5 s (10 Dateneinträge) auf 11,3 s (30 Dateneinträge). Im weiteren Verlauf ist die Annäherung an einen Grenzwert zu beobachten, der bei ca. 10,5 s liegt (100 Dateneinträge).

Durch das Sortieren der Einträge, beispielsweise nach einzelnen Wochentagen, wird die Anzahl an zur Verfügung stehenden Daten schrittweise reduziert. Um für weitere Analysen möglichst viele Zugnummer-Station-Kombinationen einbeziehen zu können, wird eine Mindestanzahl von 30 Einträgen

festgelegt. Wenngleich eine höhere Datenanzahl genauere Ergebnisse liefern würde, kann bei dieser Anzahl, basierend auf der obigen Auswertung, davon ausgegangen werden, dass die Differenz zum Grenzwert bei unter einer Sekunde liegt.

Die Analyse von Subdatensätzen ist für insgesamt 150 Kombinationen aus Zugnummer und Station in Bild 4 dargestellt. Je feiner die Unterteilung der Subdatensätze ist, desto geringer ist der durchschnittliche Fehler. Bei 50% der betrachteten Kombinationen liegt der MAE beim Berücksichtigen aller Tage von Montag bis Sonntag exkl. Feiertage (Kategorie 1) bei unter 6,1 s und für die Tage von Montag bis Freitag exkl. Feiertage (Kategorie 2) bei 5,9 s. Wird der Datensatz feiner zugeschnitten auf alle Einträge zu einem Werktag, in diesem Fall exemplarisch für Donnerstag (Kategorie 1), so sinkt der entsprechende Wert auf ca. 5,1 s.

Bei der Betrachtung einzelner Kombinationen aus Station und Zugnummer fällt jedoch auf, dass unabhängig von der Wahl des Subdatensatzes Ausreißer auftreten, bei denen die zugehörigen MAE bei bis zu 77 s liegen. Um Ursachen für solche Ausreißer zu finden, wurden weitere Datenquellen, wie bspw. Informationen zur Wetterlage

Je feiner die Unterteilung der Subdatensätze ist, desto geringer ist der durchschnittliche Fehler.

hinzugezogen. Es konnte jedoch kein signifikanter Zusammenhang zwischen Niederschlags-/Temperaturwerten und Fahrgastwechselzeiten festgestellt werden. Bei der Berücksichtigung von Zugfahrten über den Regionalverkehr hinaus wurde deutlich, dass bei einem Teil der „Ausreißerstationen“ vermehrt Zugüberholungen durch den Fernverkehr stattfinden. Die verlängerten Haltezeiten sind somit vermutlich weniger auf das Fahrgastverhalten, sondern vielmehr auf Dispositionsmaßnahmen zu Zugfolgeänderungen zurückzuführen.

Aus Bild 5 kann entnommen werden, wie stark der Einfluss von Prognosen auf den Betrieb ist. Dabei ist für einen beispielhaften Streckenabschnitt gezeigt, ab welcher Geschwindigkeit ein Ausrollen vom FAS empfohlen wird. Liegt die Prognose exakt bei der gemäß Fahrplan kalkulierten Haltezeit von 60 s, so beschleunigt der Zug nicht auf die zulässige Maximalgeschwindigkeit von 160 km/h, sondern erhält bereits vorab die FE zum Ausrollen. Dadurch erreicht er lediglich eine Geschwindigkeit von ca. 138 km/h.

Für die angefahrene Station wurden unterschiedliche Haltezeiten prognostiziert. Je kürzer die prognostizierte Haltezeit ist, desto höher ist die zur Verfügung stehende Fahrzeit auf dem Abschnitt. Dadurch ist es möglich, früher mit dem Ausrollen zu beginnen, folglich mit geringerer Geschwindigkeit zu fahren und den Energiebedarf zu verringern. Im Gegensatz dazu ist bei einer länger prognostizierten Haltezeit das Fahren mit hoher Geschwindigkeit erforderlich, um einem Verspätungsaufbau entgegenzuwirken. Ab Haltezeitprognosen von 80s wurde kein Ausrollen des Zuges empfohlen und es traten erste Verspätungen auf.

Diskussion

In diesem Artikel wurde eine Methodik zur Prognose von Fahrgastwechselzeiten mithilfe von Live-Daten vorgestellt. Neben Informationen zur Zugnummer, der angefahrenen Station und der geplanten Haltezeit wurden Daten zu der Anzahl an ein- und ausgestiegenen Fahrgästen und zur Türfreigabezeit in die Regression einbezogen. Im Zuge der Analyse wurden insgesamt drei Fragestellungen näher betrachtet. So konnte der Einfluss der Datenanzahl auf die Genauigkeit der Prognose aufgezeigt werden. Weiterhin wurden Regressionsergebnisse in Hinblick auf die Verwendung von lokalen Subdatensätzen verglichen.

Zur Einordnung der Ergebnisse wurde anschließend simulationsbasiert dargestellt, inwiefern sich unterschiedliche Prognosen auf die Generierung von FE und somit auf die Entwicklung von Verspätungen und Energieverbräuchen auswirken können.

Eine Herausforderung bei der Berechnung einer aussagekräftigen Prognose besteht in der Größe des Datensatzes. Es konnte gezeigt werden, dass sich der Prognosefehler sowohl mit steigender Datenanzahl als auch durch das Zuschneiden des globalen Datensatzes auf lokale Subdatensätze verringert. Dem entgegenstehend verringert sich die Datenanzahl beim Unterteilen der Datensätze, wodurch lediglich 150 Kombinationen aus Zugnummer und Station den definierten Mindestwert von 30 Einträgen erreichten. Um für weitere Kombinationen aus Zugnummern und Stationen robuste Regressionskurven berechnen und somit eine umfangreichere Analyse durchführen zu können, ist folglich eine größere Datenmenge erforderlich.

Eine zusätzliche Fragestellung betrifft den Umgang mit Ausreißern. Basierend auf den zur Verfügung stehenden Inputdaten konnte für einzelne Zugnummer-Station-Kombinationen keine zuverlässige Fahrgastwechselzeitprognose berechnet werden. Als Ursache hierfür werden kurzfristige Änderungen in der Zugfolge vermutet. Um die Genauigkeit der Prognosen zu optimieren,

sollten zukünftig Echtzeitinformationen zu Dispositionsmaßnahmen als weitere Datenquelle hinzugezogen werden. •

Förderhinweis

Die hier vorgestellten Arbeiten werden durch das Bundesministerium für Digitales und Verkehr (BMDV) im Programm mFUND gefördert. Die vorliegenden Inhalte sind im Rahmen des Projektes „FASaN: Fahrerassistenzsysteme adaptive Nachhaltigkeit im Bahnbetrieb“ (VKZ: 19FS2022A) entstanden.

Literatur

- [1] C. Meirich, L. Flamm, J. Krimmling und B. Jäger, „Betriebliche Optimierung des Eisenbahnbetriebs mittels vernetzter Fahrerassistenzsysteme,“ Eisenbahntechnische Rundschau (ETR), 2020.
- [2] C. Meirich, L. Flamm und B. Jäger, „Möglichkeiten und Voraussetzungen vernetzter Fahrerassistenzsysteme,“ Deine Bahn (03), pp. 48-52, 2021.
- [3] B. f. D. u. Verkehr, „Fahrerassistenzsysteme adaptive Nachhaltigkeit im Bahnbetrieb - FASaN,“ [Online]. Available: <https://bmdv.bund.de/SharedDocs/DE/Artikel/DG/mfund-projekte/fasan.html>. [Zugriff am 8. März 2024].
- [4] C. Meirich, L. Flamm, B. Jäger, P. Reiher, J. Krimmling und G. Ullrich, „Erweiterung von Fahrerassistenzsystemen im Bahnbereich durch die Verbesserung der

Datengrundlage (FAS-D),“ 2020. [Online]. Available: <https://elib.dlr.de/137684/>.

[5] D. Li, W. Daamen und R. M. Goverde, „Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station,“ Journal of Advanced Transportation, Nr. 50, pp. 877-896, 2016.

[6] P. Kecman und R. M. Goverde, „Predictive modelling of running and dwell times in railway traffic,“ Springer-Verlag, Berlin, 2015.

[7] R. A. Kuipers und C.-W. Palmqvist, „Passenger Volumes and Dwell Times for Commuter Trains: A Case Study Using Automatic Passenger Count Data in Stockholm,“ Applied Sciences, Nr. 12, 2022.

[8] scikit-learn developers, „Huber Regressor,“ [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html. [Zugriff am 11. März 2024].

Summary

Forecasting passenger changeover times with live data from operations

The further development of driver assistance systems offers the potential to optimize rail operations in terms of punctuality and energy consumption. The calculation of forecasts is becoming increasingly important in order to be able to take fluctuations in operations into account at an early stage. This article focuses on the prediction of passenger changeover times based on daily counting data.

Der Kickstart für deine Karriere!

Zahlreiche Recruiter warten an der **Talent Stage** auf dich!



90-Sekunden-Pitches | 5 Berufskategorien | 25 Kandidaten



25. September 2024 | InnoTrans Campus, Berlin







Unterstützt von:

Jetzt bewerben: www.eurailpress.de/careerboost24