

Zuverlässige Verspätungsvorhersagen mithilfe von TAROT

Christoph Stach,¹ Corinna Giebler,¹ Simone Schmidt¹

Abstract: Bei der Einhaltung von Schadstoffwerten nehmen öffentliche Verkehrsmittel eine immer entscheidendere Rolle ein. Daher wird vermehrt darauf geachtet, deren Attraktivität zu erhöhen. Ein wichtiger Punkt hierbei ist die Vorhersagegenauigkeit von Verspätungen zu verbessern, damit Fahrgäste entsprechend planen können. Die aktuell angewandten Ansätze sind häufig ungenau, da sie die zur Verfügung stehenden Daten nicht ausreichend nutzen. In diesem Beitrag stellen wir daher mit TAROT ein System vor, das mittels prädiktiver Analysen die Vorhersagegenauigkeit von Verspätungen verbessert, indem es in den Modellen Verspätungsförpflanzungen berücksichtigt. Darüber hinaus ist es in der Lage, im Fall einer Störung augenblicklich auf ein besseres Vorhersagemodell umzusteigen und auf sowohl schleichende als auch abrupte Veränderungen automatisch zu reagieren. Die Vorteile dieser Eigenschaften lassen sich in unserem TAROT-Demonstrator anhand von vier repräsentativen Anwendungsszenarien zeigen. Auch wenn sich die gezeigten Szenarien alle auf die Verspätungsvorhersage von S-Bahnen beziehen, lassen sich die Konzepte von TAROT auch auf viele andere Anwendungsbereiche (z. B. zur Bestimmung von Produktionszeiten in der Industrie 4.0) anwenden.

Keywords: Verspätungsvorhersage; ÖPNV; deskriptive Analyse; prädiktive Analyse; Concept Drift.

1 Einleitung

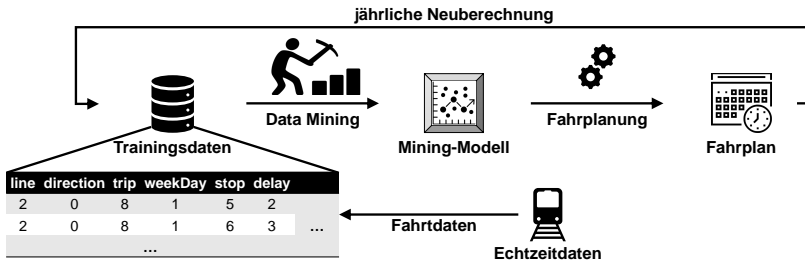
Nicht erst seit den ausgesprochenen Dieselfahrverboten sind die Bürger für das Thema „Luftreinhaltung“ sensibilisiert. Eine Möglichkeit zur Reduktion der Luftbelastung stellen öffentliche Verkehrsmittel dar. Große Verspätungen und Ausfälle erschweren den Umstieg allerdings. Auch das Bundesministerium für Verkehr und digitale Infrastruktur hat diese Problematik erkannt und sucht nach Lösungen, mit deren Hilfe Fahrplanauskünfte genauer und Reiseplanungen erleichtert werden sollen [BMVII18].

Gegenwärtig werden zwar viele Fahrtdaten erfasst, deren Potential wird bei der Fahrplanung und der Verspätungsvorhersage jedoch nicht ausgenutzt. Abb. 1(a) stellt vereinfacht dar, wie die Fahrplanung erfolgt. Fahrtdaten werden in Echtzeit erfasst und historisiert. Wurden ausreichend Daten gesammelt, werden *deskriptive Data-Mining-Techniken* darauf angewandt, um ein Modell zu erstellen, aus dem hervorgeht, wie lange Fahrten abhängig von bestimmten Einflussfaktoren dauern. Mithilfe dieses Modells kann ein Fahrplan erstellt werden. Dieser Prozess muss regelmäßig wiederholt werden, um auf veränderte Bedingungen zu reagieren.

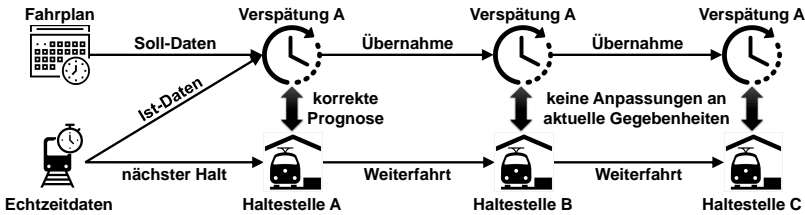
Für die Bestimmung einer Verspätung werden die so errechneten Fahrplandaten mit den aktuellen Fahrtdaten verglichen (siehe Abb. 1(b)). Die Vorhersagequalität ist für den nächsten

¹ Universität Stuttgart, Universitätsstraße 38, D-70569 Stuttgart, Vorname.Nachname@ipvs.uni-stuttgart.de





(a) Stand der Technik bei der Fahrplanung.



(b) Stand der Technik bei der Echtzeitvorhersage

Abb. 1: Stand der Technik bei der Fahrplanung und Echtzeitvorhersage.

Halt sehr gut, verschlechtert sich aber mit jedem weiteren Halt, da die Verspätung nur übernommen wird und Verspätungsfortpflanzungen nicht berücksichtigt werden [WN08]. Während die Deutsche Bahn mittlerweile bessere Konzepte für die Verspätungsvorhersage nutzt, zeigten Gespräche mit lokalen Verkehrsbetrieben, dass beispielsweise für den S-Bahn-Betrieb weiterhin diese überholten Vorhersagetechniken zum Einsatz kommen.

2 Konzept von TAROT

Um eine bessere Vorhersage zu ermöglichen, führen wir *TAROT* ein. In *TAROT* wird unter Ausnutzung aller zur Verfügung stehenden Daten ein verbessertes Vorhersagemodell errechnet, angewandt und stetig verbessert. *TAROT* basiert auf der *BRAID-Architektur* [Gi18].

Abb. 2 zeigt die Architektur von *TAROT*. Eingehende Daten werden mittels *Kafka* in zwei *Topics* (Fahr- und Störungsdaten) aufgeteilt. Die Fahrtdaten werden historisiert. Mit *Spark* werden daraus Vorhersagemodelle errechnet. Neben einem Modell für den Normalbetrieb wird für jede Störungsart (z. B., „Weichenstörung“ oder „Unfall“) ein eigenes Modell erstellt und in einem Modellspeicher abgelegt. Parallel zur Bestimmung der Soll-Daten werden die Echtzeitfahrtdaten als Ist-Daten an eine *Spark-Streaming-Komponente* weitergeleitet. Hierin ruft das *Spark-MLlib-Modul* das aktuell benötigte Modell ab und erstellt eine Verspätungsvorhersage. Sollte eine Störung eintreten, wird automatisch ein anderes Modell geladen. Die eingesetzten Vorhersagealgorithmen sind in der Lage, das Modell an *Concept Drifts* anzupassen. Neben der Vorhersage findet in *TAROT* auch eine Qualitätskontrolle

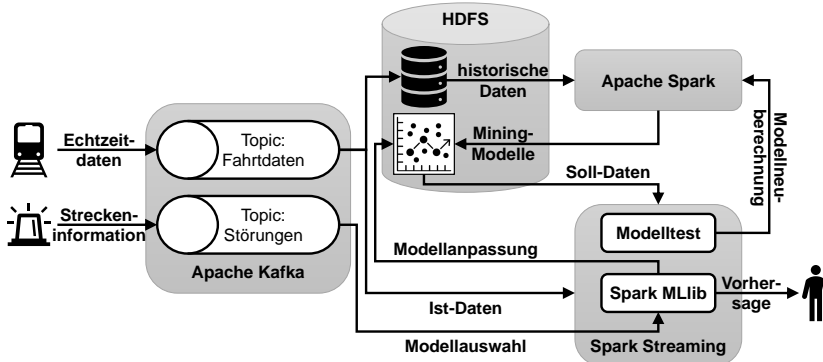


Abb. 2: Architektur von TAROT.

der Modelle statt. Sollte die Vorhersagequalität aufgrund einer abrupten Veränderung unter einen Grenzwert fallen, stößt TAROT automatisch eine vollständige Modellneuberechnung auf den historischen Daten, die seit der Veränderung gesammelt wurden, an.

3 Der TAROT-Demonstrator

Die Funktionsweise von TAROT kann anhand von vier Szenarien veranschaulicht werden. Die Szenarien verwenden jeweils synthetische Daten. Die zum Einsatz kommenden Datenmodelle orientieren sich an dem realen Datenmodell eines lokalen Verkehrsbetriebs. Es wurden allerdings irrelevante Attribute entfernt, um die Übersichtlichkeit zu erhöhen.

Demo-Szenario 1: Normalbetrieb. In dem ersten Szenario sollen die Grundfunktionalitäten von TAROT demonstriert werden. Hierfür wird auf historischen Daten ein Vorhersagemodell errechnet. Dieses Modell wird anschließend im Modellspeicher abgelegt und der Streaming-Komponente zur Verfügung gestellt. Diese ruft das Modell ab und wendet es auf eingehende Fahrdaten an. Dadurch lassen sich Verspätungsvorhersagen treffen.

In diesem Szenario wird davon ausgegangen, dass es zu keinen außergewöhnlichen Störungen kommt. Es wird zunächst lediglich der Normalbetrieb simuliert. Das heißt, es ergeben sich ausschließlich kleinere, betriebsbedingte Verzögerungen. Im Gegensatz zu der aktuell genutzten Technik zur Bestimmung der zu erwartenden Verspätung, wird in TAROT allerdings die Verspätungsförpflanzung berücksichtigt. Hierdurch wird die Vorhersagegenauigkeit wesentlich verbessert, was sich insbesondere dann auswirkt, wenn zwischen der aktuellen Position und dem Ziel mehrere Haltestellen liegen.

Demo-Szenario 2: Reaktion auf Störungen. Wird über den Streckeninformationsdienst eine Störung gemeldet, ändert sich das Fahrverhalten aller Züge schlagartig. Dies wirkt sich augenblicklich auf den Betriebsablauf aus und verändert alle Fahrzeiten. TAROT adressiert dieses Problem dadurch, dass es für jede Störungsart ein eigenes Modell gibt.

In dem zweiten Szenario wird eine dieser Störungsarten gemeldet. Die Streaming-Komponente ruft daraufhin automatisch das passende Vorhersagemodell aus dem Modellspeicher ab und wendet das neue Modell an. Dadurch sind augenblicklich alle Vorhersagen an die veränderte Situation angepasst. In dem Demonstrator wird gezeigt, welche Verspätung mit dem alten Modell vorhergesagt worden wäre, und diese wird der des angepassten Modells gegenübergestellt. Dadurch wird der Vorteil des TAROT-Vorgehens verdeutlicht.

Demo-Szenario 3: Reaktion auf schleichende Veränderungen. Neben Störungen, die augenblicklich einen Effekt auf den Betriebsablauf haben, gibt es allerdings auch kleinere Veränderungen, die sich langsam, aber stetig auf die Fahrzeiten auswirken. So kann sich beispielsweise ein Defekt an der Steckleiste mit der Zeit verschlechtern, was dazu führt, dass Züge diesen Streckenabschnitt immer langsamer durchfahren müssen.

In dem dritten Szenario wird ein Datenstrom-basierter Vorhersagealgorithmus in der Streaming-Komponente eingesetzt, der Concept Drifts erkennen und das Vorhersagemodell daraufhin anpassen kann. Das angepasste Modell wird nicht nur in der Streaming-Komponente verwendet, sondern ebenfalls im Modellspeicher abgelegt. In dem Demonstrator wird gezeigt, wie diese kleineren Anpassungen die Vorhersagequalität langfristig verbessern.

Demo-Szenario 4: Reaktion auf schlechte Genauigkeit. Trotz dieser dauerhaften Anpassungen an den Vorhersagemodellen kann es dazu kommen, dass die Vorhersagequalität zu schlecht wird, z. B. wenn ein Concept Drift abrupt sehr große Veränderungen herbeiführt. Eine Anpassung, wie sie in Demo-Szenario 3 beschrieben ist, würde zu lange zu falschen Vorhersagen führen und wäre im Fall einer abrupten Veränderung daher zu langsam.

Die Akkuratessse der Modelle kann von der Streaming-Komponente automatisch überprüft werden, sobald ein Zug am Zielbahnhof eintrifft. Ist diese zu gering, wird eine Neuberechnung des Modells angestoßen. Hierbei werden nur historische Daten berücksichtigt, die seit dem Auftreten des Concept Drifts angefallen sind. Dies wird im vierten Szenario demonstriert.

Danksagung. Die in diesem Beitrag vorgestellte Forschungsarbeit entstand aus dem PATRON-Forschungsauftrag, der von der Baden-Württemberg Stiftung finanziert wurde.

Literatur

- [BMVI18] Bundesministerium für Verkehr und digitale Infrastruktur: Digitalisierung kommunaler Verkehrssysteme, Förderrichtlinie, BMVI, 18. Jan. 2018.
- [Gi18] Giebler, C.; Stach, C.; Schwarz, H.; Mitschang, B.: BRAID — A Hybrid Processing Architecture for Big Data. In: Proceedings of the 7th International Conference on Data Science, Technology and Applications. S. 294–301, 2018.
- [WN08] Wendler, E.; Nießen, N.: Stochastische Modelle in der Eisenbahnbetriebswissenschaft. ZEVrail 132/1/2, S. 40–48, 2008.