

Automatisierte Erstellung von fahrplanreferenzierten ÖV-Netzen

Patrick Brosi und Uli Müller

geOps e.K., Freiburg · patrick.brosi@geops.de

Full paper double blind review

Abstract

Die Erstellung von fahrplanreferenzierten Linienverläufen des öffentlichen Nahverkehrs ist anspruchsvoll und wird meist, wenn überhaupt, von Hand durchgeführt. Automatisierte Ansätze scheitern meist an der Unschärfe der zugrunde liegenden Referenz-Geodaten oder an der Etablierung einer Korrelation dieser zum Fahrplan. Methoden, die auf Ähnlichkeitsmaßen entweder zwischen den Attributen oder den Geometrien (oder beidem) beruhen, setzen meist eine gewissenhafte Nachführung und eine prinzipielle Korrelierbarkeit der Attribute (über eindeutige IDs, Stations- und Liniennamen o. Ä.) in beiden Datensätzen voraus.

In dieser Arbeit präsentieren wir einen routenfindungsbasierten Ansatz, um Fahrplanverläufe und (evtl. generalisierte) Geodaten aufeinander zu korrelieren. Zur Herstellung einer Beziehung zwischen Fahrplanfahrt und Referenzgeometrie fügen wir fahrplanmäßige Stationen in die Referenzdaten ein, überführen diese in einen Graph und versuchen mittels Heuristiken insbesondere an Haltestellen und bei Lücken im Geodatensatz die topologische Korrektheit desselben zu garantieren. Wir erzeugen Fahrtverlaufsgeometrien, indem wir eine iterative Shortest-Path-Suche zwischen aufeinanderfolgenden Stationen ausführen. Die Kostenfunktion für den Wegfindungsalgorithmus ist dynamisch und folgt Heuristiken, die das optimale Resultat für das verwendete Verkehrsmittel generieren sollen.

1 Einleitung

Fahrplanreferenzierte Netze sind raumbezogene Linienverläufe die ihrer fahrplanmäßigen Entsprechung eindeutig zuordnungsbar sind. Die Anwendungsfälle solcher Netze sind zahlreich. Neben der automatisierten Erstellung von Liniennetzplänen (AVELAR et al. 2000, STOTT et al. 2010) oder der statischen Visualisierung von Fahrplänen (JOYCE 2011) können die Daten genutzt werden, um zum Beispiel Anfragen zum Geoverlauf von bestimmten Fahrten (zwischen bestimmten Stationen oder Zeitpunkten) oder komplette Geoverläufe für bestimmte Fahrten und/oder Linien auszugeben. Nutzer von Fahrplanauskunftssystemen erwarten nicht mehr nur eine klassische Routenfindung, sondern auch eine exakte georäumliche Darstellung der gewählten Verbindung. Ohne exakte, mit dem Fahrplan verknüpfte Geodaten sind solche Dienstleistungen nicht oder nur schwer anzubieten. Die Daten können außerdem dazu verwendet werden, die aktuelle Position eines Fahrzeuges aus Fahrt- und Geoverlauf und der aktuellen Verspätung des Fahrzeuges bei entsprechender Stationsdichte relativ genau durch zeitliche und räumliche Interpolation zu berechnen (BROSI 2014).

Verkehrsunternehmen verfügen heutzutage in aller Regel über einen georeferenzierten Datensatz ihrer Stationen. Zusammen mit den Fahrplaninformationen lassen sich daraus stationsgenaue Fahrtverlaufsgeometrien erzeugen. Vor allem bei großen Stationsdistanzen wirken diese Geometrien jedoch sehr rudimentär und ungenau. Häufig ist die Nachführung der Liniengeometrien zwischen den Stationen Handarbeit. Unser Ansatz erzeugt gleis- oder straßenexakte Linienverläufe automatisiert und mit beliebigen Referenzdaten.

Im Folgenden nutzen wir den Begriff „Verlauf“ für zwei verschiedene Dinge. Der „Fahrtverlauf“ ist die Abfolge der mit Zeitstempeln versehenen Stationspunkte. Der „Geoverlauf“ einer Fahrt ist der exakte Weg, den das Fahrzeug durch die Referenzdaten (und damit z. B. durch das Straßen- oder Schienennetz) nimmt. Ziel dieser Arbeit ist somit das Finden von Geoverläufen aus Fahrtverläufen.

1.1 Fahrplanreferenzierte Netze

Nach unserem Verständnis existieren drei Arten von ÖV-Netzen: der einfachste Fall, das nicht fahrplanreferenzierte Basisnetz, ist die Sammlung sämtlicher Streckengeometrien einer MOT (*Method of Transportation*), eines Verkehrsunternehmens oder einer bestimmten Region, nicht notwendigerweise versehen mit Attributen oder Hinweisen auf die Netztopologie. Das Basisnetz ist meist bei den ÖV-Betrieben vorhanden und kann auch problemlos z. B. aus OpenStreetMap extrahiert werden. Unter einem implizit fahrplanreferenzierten Netz verstehen wir das Basisnetz, das um eine Referenzierung bezüglich der Linien oder Fahrten, die auf einem Streckenabschnitt verkehren, erweitert wurde. Dabei kann ein Streckenabschnitt von mehreren (meist hunderten) Fahrten genutzt werden. Dieses Netz kann als Graph verstanden werden, bei dem die Knoten die Stationen und die Kanten die Netzsegmente darstellen und somit auch die Stationstopologie extrahierbar ist. Das explizit fahrplanreferenzierte Netz ist eine Sammlung von zusammenhängenden Geoverläufen (also nicht von Netzsegmenten) die von einer oder mehreren Fahrt(en) verwendet wird. Da das implizit und das explizit referenzierte Netz jeweils ineinander überführbar sind (entweder durch Zusammenfassung von Netzsegmenten mit gleichen Fahrt- resp. Linienreferenzierungen oder durch Segmentierung der Gesamtgeometrien mittels der Stationen als Ankerpunkte) behandeln wir sie in dieser Arbeit als äquivalent.

Fahrplanreferenzierung bedeutet hier also, dass zu jeder im Fahrplan vorkommenden Fahrt eine Liniengeometrie ausgeben werden kann und umgekehrt zu jeder Geometrie die Fahrten genannt werden können, die darauf verkehren. Dies schließt ein, dass auch die im Fahrplan vorkommenden Stationen mit ihrer topologischen Position im Netz in den Geodaten vorhanden sind.

In OpenStreetMap findet die Fahrplanreferenzierung implizit statt, indem Netzsegmenten (*ways*) eine *relation* (die Linie) zugeordnet wird (vgl. dazu [OPENSTREETMAP WIKI](#)). Diese zum Großteil händisch nachgeführte Referenzierung ist jedoch zum einen relativ ungenau (es existiert im besten Fall nur eine Linienreferenzierung, niemals eine Fahrtenreferenzierung), wird häufig bei Fahrplanänderungen erst mit langer Verzögerung oder gar nicht aktualisiert und weist außerdem v. a. im ländlichen Raum teils große Lücken auf. Einzelne, von der „Hauptachse“ einer Linie abweichende Fahrten (z. B. Rückführungen oder Verlängerungen zu Stoßzeiten) sind aus dieser Referenzierung außerdem oft nicht abzulesen.

1.2 General Transit Feed Specification

Es gibt zahlreiche Arten, Fahrplandaten maschinell zu halten (vgl. MÜLLER-HANNEMANN et al. 2007). Im deutschsprachigen Raum haben sich die Formate der EFA¹ und HAFAS² Fahrplanauskunftssysteme durchgesetzt. Ein Versuch, die Vorzüge verschiedener Fahrplandatenformate zu vereinen und diese auf einen gemeinsamen Nenner zu bringen stellt die ursprünglich von Google entwickelte General Transit Feed Specification dar (vgl. GOOGLE DEVELOPERS 2011). Da dieses Format den Fahrplan in ein sinnvolles und auch von Menschen lesbares Schema fasst, verwenden wir es hier beispielhaft, um das Vorgehen zu erläutern. Die von uns entwickelte und auf dem in dieser Arbeit beschriebene Ansatz aufbauende Anwendung ROUTEFINDER akzeptiert ebenfalls GTFS als Eingangsformat für Fahrplandaten.

GTFS besteht aus mehreren CSV-Dateien. Die wichtigsten sind:

- `trips.txt`: Die Liste der Fahrten.
- `stops.txt`: Sämtliche Haltestellen, mit WGS84 Koordinaten.
- `stop_times.txt`: Die Liste der Ankunfts- und Abfahrtszeiten für sämtliche Fahrten und Stationen.
- `routes.txt`: Routen gruppieren Fahrten z. B. einer Linie.
- `shapes.txt`: Exakte Geoverläufe von Fahrten (optional).

Fehlt die Datei `shapes.txt`, so stellen die Fahrplandaten eine Sammlung von stationsgenauen Geometrien der einzelnen Fahrten dar.

2 Netzgenerierung durch Routenfindung

Die Netzfindung kann als Korrelation des (hochgeneralisierten) stationsgenauen Netzes aus den Fahrplandaten auf Referenzdaten, z. B. OpenStreetMap oder proprietäre Quellen verstanden werden. Experimente, z. B. die Geometrien von Fernzugs-Relationen mit den Station-zu-Station-Geometrien aus Fahrplandaten über Ähnlichkeitsmaße (z. B. Fréchet-Distanz) zu vergleichen, scheitern vor allem bei großen Distanzen und/oder komplizierten Fahrtverläufen. SCHÄFERS et al. (2014) beschreiben einen Ansatz, um verschiedene Geodaten basierend auf Attributs-Ähnlichkeiten zu korrelieren. Unser Ziel war es, unabhängig von der Qualität der Referenzdaten vernünftige Ergebnisse zu erreichen.

2.1 Überführung des Referenznetzes in Graph

In einem ersten Schritt überführen wir die geografischen Referenzdaten in einen Graphen und korrigieren eventuell vorhandene Lücken. Hierzu gehört im Falle von OpenStreetMap z. B. das Suchen nach sehr nahe beieinanderliegenden, verwaisten Weg-Enden, die vor allem im Schienennetz häufig ein Zeichen für eine fehlerhafte Erfassung sind.

¹ <https://www.mentzdv.de/index.php?id=114>

² <http://www.hacon.de/hafas>

Danach werden die Stationen aus dem Fahrplan in den erzeugten Graph einsortiert, was auf verschiedene Weise geschehen kann. Da wir uns in dieser Arbeit auf GTFS als Eingangsformat beschränken, gehen wir jedoch von einer Georeferenzierung der Stationen aus. Das Einfügen der Stationen in den Graph erfolgt im einfachsten Fall durch eine Suche nach der am nächsten liegenden Kante. Heuristiken, die das Ergebnis deutlich verbessern, sind z. B. ein Snapping der Stationen auf bereits erfasste Stationen in den Geo-Referenzdaten.

Hauptproblem bei der Einfügung von im Grunde „schematischen“ Fahrplanstationen in ein konkretes, die Realität abbildendes Netz ist der Erhalt der Netztopologie. Dies ist nicht trivial, wie Abbildung 1 zeigt.

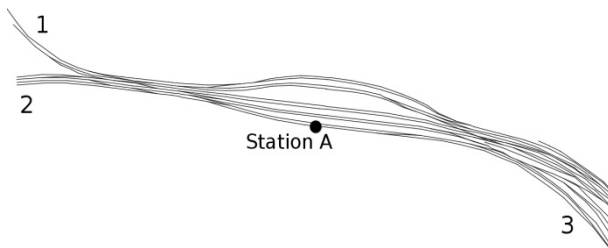


Abb. 1:
Kein topologisch korrektes Snapping – von Strang 1 ist Station A nicht zu erreichen

Station A ist korrekt auf die nächste Liniengeometrie gesnappet. Dennoch ist auf diesem Graphen Station A von Gleisstrang 1 nicht erreichbar, da von Strang 1 keine Verbindung zum südlichsten Gleis der Station besteht – die Netztopologie ist nicht mehr korrekt.

Abhilfe schaffen sog. Sekundärstationen, die auf Kanten in nächster Umgebung der Primärstation projiziert werden. Eine topologisch korrekte Einsortierung von Station A ist in Abbildung 2 zu sehen.

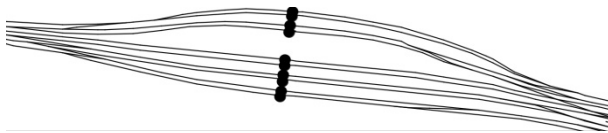


Abb. 2:
Topologisch korrektes Snapping durch Einfügen von Sekundärstationen

Die Sekundärstationen werden über Spezialkanten verbunden. Effektiv wird damit dem Fahrzeug während des Shortest-Path-Algorithmus erlaubt, das Gleis zu wechseln. Leider führt dies auch häufig zu „springenden“ Artefakten innerhalb von großen Bahnhöfen.

2.2 Korrelierung durch iterative Shortest-Path-Findung

Die eigentliche Korrelierung erfolgt, indem für jede Fahrt jeweils zwischen zwei aufeinanderfolgenden Haltepunkten ein Shortest-Path-Algorithmus ausgeführt wird (Abb. 3). In konkreten Anwendungsfällen haben wir hierfür den Algorithmus von Dijkstra verwendet (DIJKSTRA 1959).

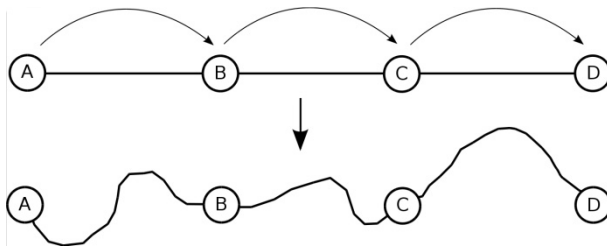


Abb. 3:
Linienverlaufsfindung durch
Shortest-Path-Iterationen

Es ist nicht garantiert, dass der kürzeste Weg auch der korrekte ist – vielmehr ist es vor allem in Schienennetzen häufig der Fall, dass der Weg mit der geringsten Distanz nicht nur „nicht korrekt“, sondern auch unmöglich von einem realen Fahrzeug zu befahren ist.

Die Menge der inkorrekten Fahrtverläufe kann durch Anwendung intelligenter Heuristiken stark reduziert werden. Heuristiken sind in diesem Fall Kostenfunktionen des Dijkstra-Algorithmus und müssen auf das Verkehrsmittel, für das Fahrtverläufe gesucht werden, angepasst werden. So hat es sich zum Beispiel als nützlich herausgestellt, Zügen das Durchfahren eines Bahnhofes ohne Halt stark zu bestrafen oder das Wenden auf freier Strecke gänzlich zu untersagen. Bei der Verwendung von ausreichend nachgeführten Referenzdaten können darüber hinaus gewisse Attribute während der Routenfindung berücksichtigt werden (triviales Beispiel: Einbahnstraßen und Straßenkategorien in OSM). Außerdem ist es über die Kostenfunktion möglich, bestimmte Wege zu begünstigen. Sind im Referenzdatensatz bereits (unvollständige) Fahrtwege erfasst, kann der von unserem Ansatz gefundene Weg durch geringere Kantenkosten auf diese Wege „gesnapt“ werden.

2.3 Plausibilitätsprüfung

Zur weiteren Qualitätssicherung werden die gefundenen Geometrien nach Abschluss der iterativen Wegfindung für eine komplette Fahrt auf Sinnhaftigkeit geprüft. Hierzu gehören unter anderem der Vergleich der Luftliniendistanz zur Distanz der gefundenen Geometrie und die Berechnung der Durchschnittsgeschwindigkeit, die ein fahrplanmäßig verkehrendes Fahrzeug auf bestimmten Abschnitten erreichen müsste.

Die Plausibilitätsprüfung findet jeweils zwischen aufeinanderfolgenden Stationen statt. Schlägt die Prüfung fehl, so wird als Fallback die direkte Station-zu-Stationen-Verbindung in die Geometrie eingefügt.

2.4 Komprimierung

Der oben beschriebene Algorithmus erzeugt in seiner naiven Form naturgemäß eine große Zahl von redundanten Geometrien. So ist es beispielsweise unnötig, bei einer Straßenbahnlinie, die im 10-Minuten-Takt verkehrt, Geometrien für sämtliche stattfindenden Fahrten zu erzeugen und zu speichern, wenn diese alle denselben Linienverlauf besitzen. Zur Vermeidung solcher Redundanzen wenden wir jeweils ein Prä- und ein Postfindungs-Äquivalenzmaß an. Fahrten, für die noch keine Geometrie gesucht wurde, können anhand ihrer Stationsfolge und ihres Fahrzeugstyps hinreichend sicher mit bereits gefundenen Geometrien verglichen und diesen zugeordnet werden. Für gefundene Geoverläufe bieten sich allgemeine Ähnlichkeitsmaße von Liniengeometrien an.

Die Qualität unseres Ansatzes hängt zu einem großen Teil von der Qualität der Referenz-Geodaten ab. Vor allem in Gebieten mit dünner Besiedelung kann es daher zu Abweichungen kommen. Die Korrektheit der gefundenen Geometrien hängt außerdem – vor allem bei Buslinien – stark von der Distanz zwischen einzelnen Stationen ab. Probleme bereiten hier z. B. Überlandbusse mit wenigen Haltestellen. Zum Zeitpunkt der Abfassung dieses Textes lagen uns keine Fahrplandaten von Fernbussen vor, sodass wir die Qualität der gefundenen Geometrien bei Bussen mit mehreren hundert Kilometern Stationsabstand noch nicht beurteilen können.

4 Fazit und Ausblick

Fahrplanreferenzierte ÖV-Netze sind für Verkehrsunternehmen unverzichtbar, wenn moderne Auskunftssysteme angeboten werden sollen. Der von uns beschriebene Ansatz wurde und wird in mehreren Anwendungsfällen auf verschiedenen, auch hochkomplexen, Referenzdaten angewandt und hat sich als zuverlässig erwiesen. Stichprobenartige Prüfungen der Netzqualität ergaben befriedigende Resultate. Eine umfassende, automatisierte Prüfung unserer Ergebnisse ist zumindest in den von uns behandelten Netzen schwierig bis unmöglich, da es an zuverlässigen, an den Fahrplan geknüpften Referenzdaten bisher noch fehlt. Für eine stichfeste Beurteilung des hier beschriebenen Ansatzes bietet es sich z. B. an, vorliegende fahrplanreferenzierte Netze (aus Shape-Dateien von GTFS-Feeds, z. B. von New York⁸) systematisch mit unseren Ergebnissen zu vergleichen.

Literatur

- AVELAR, S. & MÜLLER, M. (2000), Generating Topologically Correct Schematic Maps.
BROSI, P. (2014), Real-Time Movement Visualization of Public Transit Data.
DIJKSTRA, E. W. (1959), A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1, 269-271.
DOUGLAS, D. & PEUCKER, T. (1972), Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *The Canadian Cartographer*, 10 (2), 112-122.
GEOPS (2014), Konvertierung HAFAS-Rohdaten nach GTFS. <http://gtfs.geops.ch/doc>.
GOOGLE DEVELOPERS (2011), General Transit Feed Specification Reference. <https://developers.google.com/transit/gtfs/reference>
JOYCE, A. (2011), Automated Visualization of Public Transportation Time Schedules.
MÜLLER-HANNEMANN, M., SCHULZ, F., WAGNER, D. & ZAROLIAGIS, C. (2004), Timetable Information: Models and Algorithms. In: *Railway Optimizations*, Springer, 67-90.
OPENSTREETMAP WIKI, Overpass API. http://wiki.openstreetmap.org/wiki/Overpass_API.
OPENSTREETMAP WIKI, Public Transport Routes. http://wiki.openstreetmap.org/wiki/Relation:route#Public_transport_routes.

⁸ <http://web.mta.info/developers/download.html>

- SCHÄFERS, M. & LIPECK, U. (2014), A Similarity Based Object Matching Process for Integrating Geodatabases, In: Databases and Information Systems VIII.
- STOTT, J., RODGERS, P., MARTÍNEZ-OVANDO, J. C. & WALKER, S. G. (2010), Automatic Metro Map Layout Using Multicriteria Optimization. IEEE Transactions on Visualization and Computer Graphics, 17 (1).