

Masterthesis
Fachhochschul-Studiengang
Master Informatik

Implementierung eines Verfahrens zur automatisierten Verkehrsmittelerkennung

Transportation Mode Detection

ausgeführt von

Michael Zangerle, BSc
1310249004

zur Erlangung des akademischen Grades
Master of Science in Engineering, MSc

Dornbirn, im Juli 2015

Betreuer: Prof. (FH) DI Thomas Feilhauer

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich vorliegende Masterthesis selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Dornbirn, am 29. Juli 2015

Michael Zangerle, BSc

Zusammenfassung

Zusammen
ergänzen

Abstract

Abstract e
zen

Inhaltsverzeichnis

1	Einleitung	1
1.1	Ziele dieser Arbeit	1
1.2	Motivation und Nutzen	2
1.3	State of the art	3
1.3.1	Daten	4
1.3.2	Verkehrsmittel	4
1.4	Weiterer Aufbau der Arbeit	4
2	Hauptteil	7
2.1	Terminologie	8
2.2	Daten	8
2.2.1	Trainingsdaten	8
2.2.2	Neue Aufzeichnungen	8
2.2.3	GIS-Daten	8
2.3	Der Prototyp	8
2.3.1	Funktionalitäten	8
2.3.2	Aufbau und Architektur	8
2.4	Filter	10
2.4.1	Geschwindigkeitsfilter	10
2.4.2	Höhenfilter	10
2.4.3	Distanzfilter	10
2.5	Aufbereitung ohne GIS-Daten	10
2.5.1	Auswahl	10
2.6	Aufbereitung mit GIS-Daten	10
2.6.1	Auswahl	10
2.6.2	Abstand zu Bushaltestellen	10
2.6.3	Abstand zu Gleisen	10
2.7	Segmentierung	10
2.8	Analyse	10
2.8.1	Entscheidungsbaum	10
2.9	Nachbearbeitung	10

3	Auswertung	11
3.1	Genauigkeit ohne GIS-Daten	11
3.2	Genauigkeit mit GIS-Daten	11
4	Ausblick	13
	Literaturverzeichnis	15
	Quellcodeverzeichnis	17
	Tabellenverzeichnis	19
	Abbilungsverzeichnis	21
	Anhang	23

Todo list

Zusammenfassung ergänzen	i
Abstract ergänzen	iii
Auswertung Überblick	5

Einleitung

Eine der mittlerweile wertvollsten Ressourcen die der Mensch selbst produziert, sind Daten. Der Begriff Daten umfasst in diesem Kontext alle Informationen zur eigenen Person wie zum Beispiel Name und Adresse aber auch die eigenen Interessen, die persönliche politische Einstellungen, Informationen darüber wann man sich wo aufgehalten hat und viel mehr. Auch in der breiten Bevölkerung bekommt diese Tatsache immer mehr Wichtigkeit beigemessen wie man an Spielen wie dem "Data Dealer - Werde zum Data Dealer und sammle alle privaten Details über Freunde, Nachbarn, Bekannte und den Rest der Welt." erkennen kann [Data Dealer, 2015].

Viele dieser Daten werden sowohl bewusst als auch unbewusst mit Anderen geteilt oder auf eine andere Art und Weise in Umlauf gebracht. Das Stichwort Datenkrake ist allgegenwärtig. Ein Beispiel dafür, dass sich mit solchen Daten aber auch Ziele verfolgen lassen, ohne einzelne Personen speziell zu analysieren, soll diese Arbeit über das automatisierte Erkennen von verschiedenen Verkehrsmitteln sein.

1.1 Ziele dieser Arbeit

Das Hauptziel dieser Arbeit ist es, einen Prototypen zu erstellen, welcher anhand von aufgezeichneten GPS-Spuren und in Kombination mit verschiedenen Methodiken versucht das benutzte Verkehrsmittel mit einer möglichst hohen Wahrscheinlichkeit zu bestimmen. Diese aufgezeichneten GPS-Spuren enthalten dabei keinerlei Informationen über die jeweilige Person. Deshalb erfolgt die Auswertung ausschließlich über die GPS-Spur sowie über öffentliche zugängliche Daten (z.B. Busstationen und Gleise). Jede dieser Aufzeichnung kann mehrere Verkehrsmittel beinhalten und von unterschiedlicher Länge sein.

Es soll weiters, für jede Person die ein Smartphone besitzt möglich sein, selbst solche GPS-Spuren aufzuzeichnen und diese analysieren zu lassen. Dies bedeutet, dass keine speziellen Geräte oder andere Sensoren benötigt werden und sich dieser Prototyp mit

möglichst geringen Anpassungen auch auf andere Regionen (mit technisch weniger ausgereiften Smartphones) anwenden lassen kann. Zum Aufzeichnen der GPS-Spuren wurde für diese Arbeit die App MyTrack verwendet aber im wesentlichen funktioniert der Prototyp mit allen GPX-konformen Dateien unabhängig davon von welchem Gerät diese aufgezeichnet worden sind.

Im Zuge dieser Arbeit wird auch untersucht welches Level an Genauigkeit sowohl mit als auch ohne geografischen Informationen im Raum Vorarlberg erreicht werden kann. Dabei soll es nach der Analyse die Möglichkeit geben, die automatisch bestimmten Verkehrsmittel manuell zu korrigieren sollten diese nicht mit der Realität übereinstimmen. Auch diese manuellen Änderungen sollen in die Auswertung mit einfließen.

Schlussendlich soll auch eine Aussage über die Machbarkeit beziehungsweise den Aufwand und die möglicherweise noch zusätzlich benötigten Daten gemacht werden.

Nichtziele

In dieser Arbeit werden die diverse Schlussfolgerungsmodelle nicht betrachtet oder verglichen sondern es wird auf ein aufgrund von anderen Arbeiten vielversprechendes Modell gesetzt. Dieses Modell ist der Entscheidungsbaum. Außerdem werden die Daten zur Analyse bzw. Auswertung nicht in Echtzeit betrachtet sondern in Form einer GPS-Spur (Log) an den Prototypen übergeben.

Die Frage, wo am Körper man das Smartphone am besten trägt um genaue GPS-Daten zu erhalten sowie die Frage nach dem Energierverbrauch der App bzw. wie eine möglichst energieschonende App und die dazugehörige Kommunikation aufgebaut sein könnte bleibt unberücksichtigt. Auch eine umfassende Behandlung der Themen Sicherheit und Privatsphäre würden den Rahmen der vorliegenden Arbeit sprengen und bleibt daher unberührt.

1.2 Motivation und Nutzen

Schon früher wurde versucht Aufzeichnungen über die Verkehrswege von verschiedenen Menschen zu sammeln. Aber die Protokollen in Papierform sowie die Telefonbefragungen waren zu aufwändig und die Probanden nicht zuverlässig genug. Darum ist es von entscheidendem Vorteil eine App zur Verfügung zu haben, welche die Vorgänge des Aufzeichnens möglichst genau für einen übernimmt.

Wird eine Auswertung mit einer für das Zielgebiet aussagekräftigen Anzahl an Personen durchgeführt, so kann das Resultat für verschiedenste Zwecke verwendet werden. Auch ohne Analyse kann rein durch die Betrachtung der gesammelten GPS-Spuren festgestellt werden, welche Routen besonders häufig benutzt werden.

Zieht man nun verschiedene Werte aus der Auswertung hinzu kann auch festgestellt werden wo sich zum Beispiel verkehrstechnische Engstellen befinden und welche Routen sehr populär sind oder Aussagen über die allgemeine Verkehrssituation machen. Durch die gesammelten Daten könnten sich auch Simulationen für anstehende Bauvorhaben machen lassen und auch versucht werden eine Vorhersage für bestimmte Situationen zu tätigen. Diese Aspekte können unter anderem für das Verkehrsministerium, den öffentlichen Personennahverkehr oder auch für die Stadtplanung sehr interessant sein.

Eine ganze Reihe von Apps lässt sich mit den Auswertungen erstellen. Diese Apps könnten die Auswertungen in soziale Medien zu integrieren, für Fitnessanalysen verwendet werden, einen einfachen Rückblick über die eigene Fortbewegung ermöglichen oder für Umweltbewusste errechnen wie viel CO₂ sie produziert oder gespart haben. Ein Reisetagebuch könnte daraus genau so Nutzen ziehen wie eine App die beim Autofahren Auskunft über die aktuell billigste Tankstelle in näherer Umgebung gibt oder eine App die einfach nur Vorschläge für alternative Routen zu einem bekannten Ziel anbietet.

Zusammenfassend kann man sagen, dass die Verwendungsmöglichkeiten für solche Daten umfangreich sind und sich am besten unter den Begriffen kontextorientierte, geographische Apps zusammenfassen lassen. Nicht zuletzt öffnen sich mit solchen Daten aber auch umfangreiche Möglichkeiten für die Werbebranche.

1.3 State of the art

Zu den Meilensteinen auf diesem Gebiet zählt sicher die Arbeit von Yu Zheng in welcher er unter anderem auf die Erkennung von den Abschnitten mit nur einem Verkehrsmittel eingeht. Weiters verwendete er in seiner Arbeit auch einen mit dem von den GPS-Spuren gesammelten geographischen Wissen aufgebauten Graphen welcher zur weiteren Auswertung verwendet wurde. [Zheng et al., 2010]

Mit der Frage wie geographische Daten in eine solche Analyse miteinbezogen werden können hat sich auch Leon Stenneth beschäftigt. Dabei hat er nicht nur fixe Daten wie Gleise und Busstationen sondern auch aktuelle Buspositionen miteinbezogen. [Stenneth et al., 2011]

Sowohl Stenneth als auch Zheng haben in ihren Arbeiten detailliert erklärt wieso sie welche Attribute (Geschwindigkeit, Beschleunigung, ...) für die Bestimmung des Verkehrsmittels verwendet haben und sie haben diese auch durch Versuche nach ihrer Wichtigkeit gereiht. Außerdem haben beide und auch Sasank Reddy mehrere Schlussfolgerungsmodelle (Entscheidungsbaum, Bayessches Netz, Markov Modelle, Random Forest, ...) betrachtet und miteinander verglichen. [Reddy et al., 2010, Stenneth et al., 2011, Zheng et al., 2010]

Wie man mit Verbindungsabbrüchen umgeht und zwischen ähnlichen Verkehrsmitteln unterscheiden kann hat unter anderem auch Filip Biljecki beschäftigt. Außerdem baut

er für die unterschiedlichen Kategorien von Verkehrsmittel ein hierarchisches Modell auf welches ihm helfen soll bessere Entscheidungen zu treffen. [Biljecki et al., 2013]

1.3.1 Daten

Ein wesentlicher Unterschied zwischen all den betrachteten Publikationen sind die verwendeten Daten. Einzig die GPS-Spuren bilden eine gemeinsame Basis. Manche untersuchten die Verwendung von GSM- und WIFI-Informationen [Reddy et al., 2010], stützten sich auf Zusatzinformationen durch weitere Sensoren wie zum Beispiel ein Beschleunigungssensor [Reddy et al., 2010, Nadine Schüssler et al., 2011]. Andere wie Leon Stenneth verwendeten Live-Informationen von den öffentlichen Verkehrsmittel und kombinierten diese mit GIS-Informationen, seien es Busstationen, Bahnstrecken, das Straßennetz oder Parkplätze [Stenneth et al., 2011].

1.3.2 Verkehrsmittel

Ein weiterer Unterschied zwischen den Publikationen sind die betrachteten Verkehrsmittel. Hierbei reicht die Spanne der unterschiedenen Verkehrsmittel von "Gehen und Motorisiert" (siehe [Reddy et al., 2010]) bis hin zu "Gehen, Zug, U-Bahn, Rad, Auto, Straßenbahn, Bus, Fähre, Segelboot und Flugzeug" (siehe [Biljecki et al., 2013]).

1.4 Weiterer Aufbau der Arbeit

Der Hauptteil der vorliegenden Arbeit gliedert sich in 4 große Abschnitte:

Im ersten Abschnitt wird eine Terminologie für die weiteren Abschnitte festgelegt um Missverständnisse und Mehrdeutigkeiten zu beseitigen. Danach wird auf die Akquirierung der GPS-Daten eingegangen. Dies umfasst sowohl die gesammelt GPS-Spuren und deren Struktur, sowie die verwendeten GIS-Daten. Dabei geht es einerseits um deren Herkunft als auch darum wie diese extrahiert wurden und wie auch diese Daten aufgebaut sind. Schlussendlich wird ein Überblick über den entstandenen Prototypen sowie dessen Handhabung und Architektur vermittelt.

Der zweite Abschnitt befasst sich sowohl mit dem Filtern als auch dem Aufbereiten der Daten sowie das Aufteilen von GPS-Spuren in Teile in denen nur ein Verkehrsmittel verwendet wurde. Filtern der Daten bedeutet in diesem Zusammenhang, dass Außreiser aus den GSP-Spuren entfernt werden. Diese Ausreißer können aufgrund von Geschwindigkeitssprüngen als auch unwahrscheinlich große Sprünge im dreidimensionalen Raum sein.

Mit Aufbereiten der Daten ist gemeint, dass Werte zu GPS-Punkten für die spätere Analyse berechnet werden. Diese Werte können sowohl die Geschwindigkeit, Beschleunigung, Distanz, Zeit als auch der Abstand zu der nächsten Bushaltestelle sein. Schlussendlich sollen die gefilterten und erweiterten Daten verwendet werden um eine einfache und sichere Aufteilung der GPS-Spuren anhand der Verkehrsmittel zu ermöglichen.

Aufbauen auf die Resultate aus dem zweiten Abschnitt befasst sich der dritte Abschnitt mit der tatsächlichen Erkennung von der Verkehrsmittel anhand der berechneten Werte und den einzelnen Abschnitten der GPS-Spur. Dabei wird auch genauer auf die verwendete Schlussfolgerungsmethode den Entscheidungsbaum sowie dessen Training eingegangen. Die aus dem Entscheidungsbaum gewonnenen Erkenntnisse werden schlussendlich ein letztes Mal überprüft um z.B. sinnfreie Wechsel zwischen Verkehrsmitteln zu verhindern.

Im vierten und letzten Abschnitt des Hauptteils befasst sich mit der Auswertung der gewonnenen Erkenntnisse aus den vorhergehenden Abschnitten sowie den Testläufen mit neuen GPS-Spuren.

Auswertung
Überblick

Hauptteil

2.1 Terminologie

2.2 Daten

2.2.1 Trainingsdaten

Akquirierung

Struktur

2.2.2 Neue Aufzeichnungen

Akquirierung

Struktur

2.2.3 GIS-Daten

Akquirierung

Struktur

2.3 Der Prototyp

2.3.1 Funktionalitäten

2.3.2 Aufbau und Architektur

2.4 Filter

2.4.1 Geschwindigkeitsfilter

2.4.2 Höhenfilter

2.4.3 Distanzfilter

2.5 Aufbereitung ohne GIS-Daten

2.5.1 Auswahl

Durchschnittliche Geschwindigkeit

Maximale Geschwindigkeit

Durchschnittliche Beschleunigung

Maximale Beschleunigung

Höhenmeter

Distanz

2.6 Aufbereitung mit GIS-Daten

2.6.1 Auswahl

Durchschnittliche Geschwindigkeit

Maximale Geschwindigkeit

Durchschnittliche Beschleunigung

Maximale Beschleunigung

2.6.2 Abstand zu Bushaltestellen

2.6.3 Abstand zu Gleisen

2.7 Segmentierung

2.8 Analyse

Auswertung

3.1 Genauigkeit ohne GIS-Daten

3.2 Genauigkeit mit GIS-Daten

Ausblick

Literaturverzeichnis

- [Biljecki et al., 2013] Biljecki, F., Ledoux, H., and van Oosterom, P. (2013). Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27(2):385–407.
- [Data Dealer, 2015] Data Dealer (2015). Data dealer. legal, illegal, scheissegal! <http://datadealer.com/de>.
- [Nadine Schüssler et al., 2011] Nadine Schüssler, Lara Montini, and Christoph Dobler (2011). Improving post-processing routines for gps observations using prompted-recall data. In *9th International conference on survey methods in transport*.
- [Reddy et al., 2010] Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13.
- [Stenneth et al., 2011] Stenneth, L., Wolfson, O., Yu, P. S., and Xu, B. (2011). Transportation mode detection using mobile phones and GIS information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 54–63. ACM.
- [Zheng et al., 2010] Zheng, Y., Chen, Y., Li, Q., Xie, X., and Ma, W.-Y. (2010). Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web (TWEB)*, 4(1):1.

Listings

Tabellenverzeichnis

Abbildungsverzeichnis

Anhang