

26. Verkehrswissenschaftliche Tage 2018

Ein datengetriebener Predictive-Modeling-Ansatz zur Evaluation potentieller Bikesharing-Entleihstationen

Mathias Willnat¹, Dr. Martin Treiber²

² TU Dresden, Institut für Wirtschaft und Verkehr
Lehrstuhl für Ökonometrie und Statistik, insb. im Verkehrswesen, 01062 Dresden

1 Einleitung

Unter Bikesharing versteht man Mobilitätsangebote, die Privatkunden eine kurzfristige Entleihe einzelner Fahrräder insbesondere für Punkt-zu-Punkt-Verbindungen ermöglichen. Aus betriebswirtschaftlicher Sicht sollten die Standorte der Entleihstationen so gewählt werden, dass der erwartbare Gewinn im gesamten System maximiert wird. Konkret bedeutet dies, dass bei der Evaluation eines potentiellen zusätzlichen Standorts nicht nur die Nachfrage vor Ort zu schätzen ist. Ebenso sollten die induzierten Netzwerkeffekte im Bestandsnetz berücksichtigt werden. Drittens sind die hohen Kosten der Redistributionstransporte zu betrachten, die bei starken zeitlichen Dysbalancen zwischen Entleihen und Rückgaben an einer Station notwendig werden, um die Verfügbarkeit von Rädern an allen Stationen sicherzustellen.

[1-4] liefern Beispiele für Regressionsanalysen, mithilfe derer mittlere Nutzungshäufigkeiten anhand von raumstrukturellen Variablen wie Bahnhofsdistanz oder Bebauungsdichte geschätzt werden. [5-7] erweitern diese Ansätze mittels Mixed-Effects-Regressionen um zeitpunktabhängige Kennzahlen wie Temperatur oder Niederschlagsmenge. Weitgehend unbeachtet bleiben in diesen Arbeiten jedoch Netzeffekte der untersuchten Bikesharingssysteme. Eine valide Schätzung der künftigen Radnachfrage an einer neuen Station muss beispielsweise auch Kannibalisierung- und Erreichbarkeits-effekte im gesamten Netz berücksichtigen. Für eine Evaluation potentieller neuer Entleihpunkte scheinen diese Modelle somit kaum geeignet.

Daneben existieren verschiedene Modelle mit dem expliziten Ziel der Bikesharing-Stationsnetzplanung, vgl. bspw. [8-10]. Dabei werden vornehmlich mathematische Algorithmen der linearen Optimierung unter komplexen Systemen von Nebenbedingungen verwendet. Zwar zielen diese Ansätze ausdrücklich auf die bestmögliche Berücksichtigung der Netzeffekte, meist dienen hier jedoch lediglich allgemeine Verkehrsnachfrageerhebungen als Datengrundlage. Dies zwingt nicht nur zu restriktiven Vorannahmen, es bleiben gleichzeitig auch bikesharingtypische Verkehrsmuster gänzlich unberücksichtigt, wie beispielsweise der empirisch sehr hohe Anteil von Last-Mile-Fahrten.

¹ Korrespondierender Autor: Email: mail@mathias-willnat.de

² Email: martin.treiber@tu-dresden.de

In [11] und [12] gelingt die Einbeziehung von Netzeffekten bei Bikesharing-Nutzungsdaten mithilfe eines Nearby-Station-Scores, eines distanzbasierten Lagemaßes gegenüber anderen Stationen. Die Nachfrage wird anhand von Machine-Learning-Methoden wie Neuronalen Netzen und Ensembleverfahren prognostiziert. Gleichzeitig werden neue Stationen einer binären Balancekategorie [11] bzw. einem Nutzungscluster [12] zugewiesen. Damit liefern sie die bisher einzigen uns bekannten Planungsansätze, in denen auf Grundlage empirischer Nutzungsdaten unter Berücksichtigung von Dichteeffekten Nachfrageprognosen ermittelt werden, die zumindest ungefähre Rückschlüsse auf die Nutzungsbalance zulassen. Ein Abschätzen stationsspezifischer Tagesgänge von Entleihen und Rückgaben ist auf Grundlage von groben Balanceclustern jedoch kaum sinnvoll möglich.

In diesem Beitrag wird hierfür erstmals ein Framework vorgestellt, mithilfe dessen potentielle Standorte hinsichtlich der erwarteten absoluten Nutzungshäufigkeit an der Station selbst, der im Bestandsnetz zusätzlich induzierten Nachfrage sowie der erwartbaren Tagesgänge von Entleihen und Rückgaben evaluiert werden können. Anhand mehrerer Millionen Einzelbuchungen im Bikesharingsystem Frankfurt/Main, zahlreicher stationsgenauer Raumstrukturdaten und unter Berücksichtigung zeitlicher Einflüsse wie Wettereffekte und Saisonfiguren wird hierfür zunächst ein umfassendes Mixed-Effects-Regressionsmodell präsentiert. Dabei wird aufgezeigt, wie Kannibalisierungs-, Dichte- und Größeneffekte des Bestandsnetzes einander überlagern und wie diese je zu quantifizieren sind. Ein Scoring von rund 4000 Straßenkreuzungen im Stadtgebiet ergibt unter anderem, dass eine Verdichtung des Kernnetzes eine höhere absolute Nachfragesteigerung bedeutet, als eine verstärkte Erschließung der Peripherie. Ein System von 48 automatisiert und unüberwacht spezifizierten Regressionsgleichungen liefert zudem passgenaue Prognosen über die Tagesgänge der Entleihen und Rückgaben an Werktagen und Wochenenden, mittels derer Dysbalancen deutlich flexibler abgeleitet werden können als über Nutzungscluster [11, 12].

Dieser Beitrag ist folgendermaßen aufgebaut: Im Anschluss an diese Einleitung werden im Abschnitt 2 die verwendeten Analysedaten vorgestellt. Nachfolgend wird im Abschnitt 3 kurz das methodische Vorgehen erläutert. Im Abschnitt 4 werden die Ergebnisse für den Anwendungsfall Frankfurt/Main vorgestellt. Der abschließende Abschnitt 5 fasst abschließend wichtige Ergebnisse zusammen.

2 Datengrundlage

Der verwendete Analysedatensatz umfasst sämtliche Buchungen des stationsbasierten Bikesharingsystems *Call-a-Bike* innerhalb des Stadtgebiets Frankfurt/Main im Zeitraum vom Januar 2014 bis Mai 2017 ($N > 2,1$ Mio.), namentlich die sekundengenauen Zeitpunkte der Entleihen und Rückgaben sowie die genauen Geokoordinaten der genutzten Stationen [13]. Diese Buchungsdaten wurden durch zahlreiche zusätzliche raumstrukturelle und zeitpunktabhängige Variablen angereichert.

Hierbei wurden für jede der 306 Stationen die Höhenlage (m über NN) [14] und die Flusseite bestimmt, ebenso die Entfernung zum Stadtzentrum, zum nächstgelegenen U-, S- bzw. Regionalbahnhof, zum nächsten Hochschulstandort sowie zur nächsten großen Einkaufsstraße bzw. -zentrum (je absolut in Metern, logarithmiert und als binarisierte Dummyvariable). Ferner wurde ein binärer Hauptbahnhofsdummy bestimmt. Zusätzlich wurde mittels einer Faktorenanalyse eine Menge von insgesamt 13 verschiedenen stadtteilgenauen

Bebauungs-, Bevölkerungs- und Wirtschaftskennzahlen [15] auf die vier Faktoren Seniorenprägung, Prosperität, Urbanität und Familienprägung reduziert und diese den Stationen zugeordnet. Darüber hinaus wurde mithilfe der Bebauungspolygone von *OpenStreetMap* [16] für jede Station der Anteil überbauter Fläche im Umkreis von 50 Metern als Schätzer für die Sichtbarkeit ermittelt. Zwei Schätzer für den Bevölkerungs- und Arbeitsplatzeinzug wurden als Produkt der überbauten Fläche im Umkreis von 250 Metern mit den stadtteilüblichen Bevölkerungs- bzw. Arbeitsplatzdichten bestimmt (durchschnittliche Anzahl je überbautem m² im näheren Umkreis).

Jeder Zeitpunkt des Untersuchungszeitraums wurde zudem mit stundengenauen Wetterdaten zu Niederschlag (in mm/h und als ordinale Niederschlagskategorien), Lufttemperatur (°C), Luftfeuchtigkeit (Prozent) und Windstärke (m/s) [17] angereichert, ferner mit Variablen als Wochentag- und Feiertagsindikatoren. Aus den stundengenau aufaggregierten Entleihzahlen konnte zudem per gleitendem Mittel eine Saisonfigur bestimmt werden (einwöchige Periodizität), mittels derer sich die Nutzungskennzahlen um übliche tageszeitliche und wochenendbedingte Schwankungen bereinigen lassen.

Während des beinahe dreieinhalbjährigen Untersuchungszeitraums hat sich die Zahl der aktiven Stationen von circa 100 auf 306 erhöht. Jedem Zeitpunkt wurde deshalb zudem die Anzahl der aktiven Stationen zugeordnet, um Netzgrößeeffekte abbilden zu können.

Neben diesen rein räumlichen oder zeitabhängigen Kennzahlen jeder Station bzw. jedes Zeitpunktes wurden ferner zwei raumzeitliche Variablen berechnet. Die Nearest-Neighbour-Distanz beschreibt für jede Station zu jedem Zeitpunkt zwischen ihrer erst- und letztmaligen Nutzung (Aktivitätsspanne) die Luftliniendistanz in Metern zur nächstgelegenen ihrerseits aktiven Station. Diese Kennzahl erlaubt die Betrachtung von Kannibalisierungseffekten.

Angelehnt an das Konzept der Lagegunst [18, S. 49 ff.] wurde zudem eine Netzlageattraktivität ermittelt. Diese Größe soll die Netzdichteeffekte abbilden, indem für jede Station zu jeder Stunde die Entfernung zu jeder anderen dann aktiven Station gemessen, bewertet und zu einer Kennzahl aggregiert wird. Grundlage hierfür bildet das Logit- bzw. Wilson-Bewertungsmodell mit einer exponentiellen Bewertungsfunktion der Form

$$B_{ij} = \exp(-\zeta \cdot W_{ij}), \quad (1)$$

wobei B_{ij} die Bewertung der Relation zwischen den Punkten i und j beschreibt und sich proportional zu der Wahrscheinlichkeit verhält, dass Fahrt realisiert wird. W_{ij} steht für einen Widerstand und ζ für die zu ermittelnden Entfernungssensitivität [18, S. 54 ff.]. Dabei lässt sich $\frac{1}{-\zeta}$ als typische Entfernung interpretieren, bei der das Radfahren noch attraktiv ist, beziehungsweise, in der Interpretation als Logitmodell, als in Metern ausgedrückte Unschärfe des Zufallsnutzens (Bewertungsunschärfe).

Die Abbildung 1a zeigt die Dichte-Histogramme aller im Untersuchungszeitraum realisierten Fahrten (blau) und aller möglichen Relationen (grün) in Abhängigkeit von der Luftliniendistanz. Es ist zu erkennen, dass in einem Entfernungsbereich von bis zu 3000 Metern, in dem ein Großteil der durchgeführten Fahrten liegt, die Zahl der möglichen Relationen (also die Zielmenge) nahezu linear ansteigt. Dies erlaubt die vereinfachende Annahme, dass es sich (für einen Großteil der relevanten Fälle) um ein homogenes Beobachtungsgebiet handelt. Es

lässt zeigen, dass in diesen Fällen aus der Bewertungsfunktion der Gleichung 1 eine Dichtefunktion der Form

$$f(W) = \zeta^2 \cdot W \cdot \exp(-\zeta \cdot W) \quad (2)$$

folgt. Rot dargestellt ist in Abb. 1a die Dichtefunktion

$$f(W) = 0,0012^2 \cdot W \cdot \exp(-0,0012 \cdot W), \quad (3)$$

wobei W hier die Luftlinienentfernung in Metern beschreibt und $\zeta = 0,0012$ durch optische Anpassung an die blau dargestellte Fahrtenhäufigkeitsverteilung ermittelt wurde. Die Berechnung der hier definierten Netzlageattraktivität einer Station i erfolgt schließlich für jeden Zeitpunkt über

$$NetzAttr_i = \sum_{j \neq i} B_{ij} = \sum_{j \neq i} \exp(-\zeta \cdot D_{ij}), \quad (4)$$

wobei sich B_{ij} mit $\zeta = 0,0012$ und D_{ij} als Luftliniendistanz zwischen i und j ergibt (siehe Verlauf in Abb. 1b).

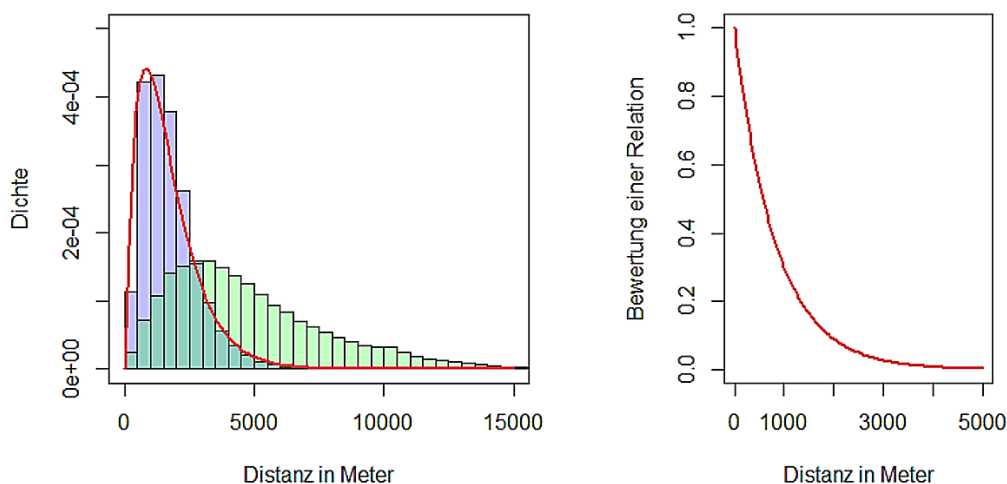


Abbildung 1a (links): Dichte aller gefahrenen Distanzen (blau), die Distanzen aller möglichen Relationen (grün) und der Verlauf der geschätzten Dichte nach Gleichung 3 (rot).

Abbildung 1b (rechts): Distanzabhängiger Verlauf der Bewertungsfunktion $B(D_{s_i s_j})$ entsprechend der Gleichung 1 bei $\zeta = 0,0012$.

Die raumstrukturellen, die zeitpunktabhängigen und die raumzeitlichen Daten wurden zu einem Paneldatensatz zusammengefügt, in dem jede Beobachtung eine Stationsstunde beschreibt. Zu jeder Stationsstunde wurde schließlich noch die Anzahl der dortigen Entleihen und Rückgaben ermittelt.

3 Methodisches Vorgehen

Zunächst soll untersucht werden, welche Einflussfaktoren sich in welcher Weise auf die absolute Nachfrage an einer Entleihstation auswirken. Wegen der geschachtelten Struktur der

Paneldaten würde ein klassisches einstufiges Regressionsmodell die Unabhängigkeitsannahme der Fehlerterme verletzen, derart gewonnene Schätzparameter wären demnach verzerrt und inkonsistent [19]. Darum wird hier stattdessen eine zweistufige Mixed-Effects-Regression der Form

$$\ln(E_{s,t} + 1) = \alpha + X\beta + \epsilon \quad (5)$$

geschätzt, wobei $E_{s,t}$ die Anzahl der Entleihen an Station s zum Zeitpunkt t beschreibt, X die Merkmalsmatrix bestehend aus einem Intercept und sämtlichen oben genannten räumlichen, zeitlichen oder raumzeitlichen Variablen, β einen Vektor zu ermittelnder Schätzparameter und ϵ einen Vektor von Fehlern. Der Vektor α steht für den stationsabhängigen Term (Random Effects). Jedes Element $\alpha_{s,t} \in \alpha$ hat die Form

$$\alpha_{s_i} = v_{0|s} + v_{1|s} \cdot \ln(NNDist_{s,t}) + v_{2|s} \cdot \ln(NetzAttr_{s,t}) + \psi_{s,t}. \quad (6)$$

Dabei steht $v_{0|s}$ für den Random Intercept jeder einzelnen Station, $v_{1|s}$ und $v_{2|s}$ für die Random Slopes der Nearest-Neighbour-Distanz $NNDist_{s,t}$ bzw. die Netzlageattraktivität $NetzAttr_{s,t}$ und $\psi_{s,t}$ für den Fehlerterm im Sublevel.

Bei der Anzahl der Entleihen einer Stationsstunde $E_{s,t}$ handelt es sich um niedrigwertige Zähldaten. Um der Normalverteilungsannahme der Fehlerterme näherungsweise gerecht zu werden, ist eine Logarithmierung der erklärten Variable nötig. Auf eine Poisson- oder negativ-binomiale Regression eines so komplexen Mixed-Effects-Modells musste hier aus Performancegründen verzichtet werden. Ebenso konnte die Modellschätzung nicht anhand sämtlicher rund 6 Millionen Stationsstunden erfolgen, stattdessen wurde eine zufällige Stichprobe von 300.000 Beobachtungen gezogen. Über einen Backward-Selection-Prozess wurden iterativ jene Variablen aus dem Modell entfernt, deren Berücksichtigung keine zusätzliche Modellgüte im Sinne eines Likelihood-Ratio-Tests bedeutet. Die so gewonnenen Fixed-Effects-Schätzparameter des Vektors β ermöglichen die Quantifizierung der mittleren Nachfrageeffekte der Kennzahlen der Merkmalsmatrix X .

Die Sublevel-Schätzer $v_{i|s}$ sind dagegen stationsindividuell. Für den Erwartungswert von α gilt $E(\alpha_s) = 0 \forall i$ [20]. Somit kann für mittlere Wetterdaten und zu mittlerer Saisonfigur mithilfe der Toplevelgleichung eine erwartungstreue Nachfrageprognose unter Berücksichtigung der Netzeffekte für jeden beliebigen potentiellen Standort getroffen werden, zu dem die Raumstrukturdaten vorliegen. Zusätzlich hätte die Hinzunahme einer neuen Station Netzeffekte auf die Nachfragen der Stationen im Bestandsnetz. Über die ermittelten Top- und Sublevelschätzer lässt sich hierzu je eine stationsindividuelle Prognose treffen.

Wie oben beschrieben verursachen starke tageszeitliche Dysbalancen von Entleihen und Rückgaben hohe Kosten für dann notwendige Redistributionstransporte. Deswegen sind neben den erwartbaren absoluten Nachfragen an den potentiellen neuen Stationen und den erwartbaren absoluten Nachfrageeffekten im Bestandsnetz zusätzlich die Tagesgänge der Fahrtbeginne und -enden an jedem zu evaluierenden Standort zu berücksichtigen.

Um dies zu modellieren wurde zunächst eine Menge von 48 zu erklärenden Variablen definiert: Unterschieden nach Werktag und Wochenende (inkl. Feiertag) wurde der Tag in 12 Zwei-Stunden-Intervalle zerlegt. Für jedes der sich so ergebenden 24 Intervalle wurden je Station

die absoluten Häufigkeiten der darauf entfallenden Entleihen ermittelt. Diese wurden anschließend derart normiert, dass je Station deren Summe Eins ergibt, wobei Wochentage mit dem Faktor Fünf und Wochenenden mit dem Faktor Zwei gewichtet wurden. Mit den Rückgaben wurde gleichermaßen verfahren. Es ergeben sich also 48 gewichtete relative Häufigkeiten.

Um robuste Ergebnisse zu erhalten wurden einige extrem geringfrequentierte Stationen dabei ausgeschlossen. Die verbleibenden 278 Stationen wurden im Verhältnis 4:1 in Trainings- und Validierungsdaten aufgeteilt. Mittels ersterer wurden anschließend 48 lineare Regressionsgleichungen geschätzt, wobei je die relativen Entleih- bzw. Rückgabehäufigkeiten als Regressanden dienten und die Menge sämtlicher in Abschnitt 2 beschriebener Raumstrukturdaten als Regressoren genutzt wurde. Für jedes dieser 48 Modelle wurde nun völlig unabhängig und unüberwacht ein Backward-Selection-Prozess angestoßen, bei dem je solange iterativ die Variablen mit dem geringsten Erklärungsgehalt ausgeschlossen wurden, bis eine weitere Modellreduktion keine Verbesserung mehr im Sinne einer AIC-Optimierung bedeuten würde.

Mithilfe dieses Systems von 48 völlig verschieden spezifizierten Modellgleichungen lassen sich Prognosen über relativen Verteilungen von Entleihen und Rückgaben an Werktagen und Wochenenden im Tagesgang treffen. Mögliche Dysbalancen lassen sich so für jeden potentiellen Standort im Untersuchungsgebiet schätzen, zu dem die Raumstrukturdaten vorliegen.

4 Ergebnisse

Der Feature-Selection-Prozess des zweistufigen Mixed-Effects-Modells ergab, dass im Toplevel neben dem Intercept auch die Regressoren Seniorenprägung, logarithmierte Bahnhofsdistanz, logarithmierte Hochschuldistanz, Zentrumsdistanz, Arbeitsplatzenzug, Höhenlage, Hauptbahnhofslage, Anzahl aktiver Stationen, die verschiedenen Wetterdaten, die Saisonfigur sowie die Nearest-Neighbour-Distanz und die Netzlageattraktivität je signifikant im Sinne eines Likelihood-Ratio-Tests sind. Die zugehörigen Toplevel-Schätzparameter sind der Tabelle 1 ersichtlich. Sie zeigen allesamt die nach Plausibilität und Forschungsstand erwartbaren Vorzeichen.

Bei der Nearest-Neighbour-Distanz und der Netzlageattraktivität ist zu beachten, dass sich die Effekte je Station als Summe des Top- und Sublevelschätzers ergeben. Auf eine Auflistung aller 918 stationsindividuellen Sublevelschätzer wird an dieser Stelle verzichtet. Es zeigt sich jedoch, dass bei 255 der 306 Stationen der so berechnete Gesamtanstieg der Nearest-Neighbour-Distanz positiv ist. An 289 Stationen war zudem ein positiver Gesamtanstieg der Netzlageattraktivität beobachtbar. Als dritte netzspezifische Variable zeigt sich auch der Toplevelschätzer der Anzahl aktiver Stationen positiv.

Diese Ergebnisse bekräftigen damit die eingangs vermuteten und teils konträren Wirkzusammenhänge: Je geringer die unmittelbare Distanz zu einer anderen aktiven Station, desto geringer ist die Nachfrage (Kannibalisierungseffekt). Gleichzeitig gilt: Je zentraler im Netz gelegen, also je mehr potentielle Zielstationen im relevanten Umkreis, desto größer ist die Nachfrage (Netzdichteeffekt). Und zusätzlich gilt: Je mehr Stationen insgesamt (also irgendwo) aktiv sind, desto mehr Räder werden auch an jeder einzelnen Station gebucht (Netzgrößeeffekt).

Tabelle 1: Schätzparameter im Toplevel (Fixed Effects) des finalen Modells der Gleichung 5

Intercept	0,2039		
Räumlich			
Seniorenprägung des Stadtteils	-0,0168	Höhenlage	-0,0006
Zentrumsdistanz	$-2,7 \cdot 10^{-6}$	log. Hochschuldistanz	-0,0195
log. Bahnhofsdistanz	-0,0161	Dummy Hauptbahnhof	0,8686
Arbeitsplatzeinzug	$4,9 \cdot 10^{-6}$		
Zeitpunktabhängig			
Lufttemperatur	0,0052	Luftfeuchtigkeit	-0,0007
Windstärke	-0,0043	Dummy Sprühregen	-0,0390
Dummy Regen	-0,0651	Dummy Starkregen	-0,0986
Saisonfigur	0,1598	Anzahl akt. Stationen	$2 \cdot 10^{-5}$
raumzeitlich (Achtung: nur Fixed Effects)			
log. Nearest-Neighbour-Distanz	0,0271	Netzlageattraktivität	0,0064

Es ist hervorzuheben, dass diese Erkenntnisse zwar hochgradig intuitiv, wegen der natürlichen negativen Korrelation von Nearest-Neighbour-Distanz und Netzlageattraktivität jedoch keineswegs trivial sind. Schließlich zählt die Distanz zur nächstgelegenen aktiven Station auch per Definition in genau entgegengesetzter Wirkrichtung am höchsten auf die Netzlageattraktivität ein.

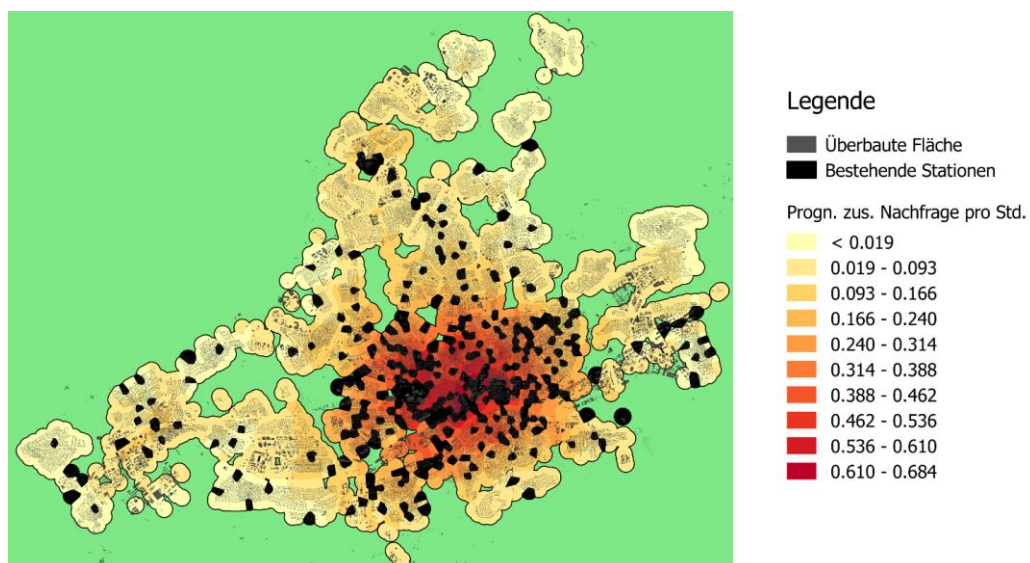


Abbildung 2: Prognostizierte Nachfrageänderung (im Mittel pro Stunde) bei Aktivierung zusätzlicher Station: Summe aus Nachfrage am neuen Standort und Netzeffekten im Bestandsnetz

Ein Scoring jener circa 4000 Straßenkreuzungen im Stadtgebiet mit einem Mindestabstand von 150 Metern zur nächstgelegenen aktiven Station (unter mittlerer Saisonfigur, mittleren Wetterlagen, und dem Bestandsnetz wie zum Ende des Beobachtungszeitraums) liefert die in Abbildung 2 abzulesenden prognostizierten zusätzlichen absoluten Nachfragen. Dabei wurden die erwarteten Nachfragen an den potentiellen Stationen selbst, sowie deren erwartete induzierte Nachfrageeffekte im Bestandsnetz berechnet und aufsummiert. Es ist klar erkennbar, dass eine Verdichtung des Kernnetzes eine höhere zusätzliche Nachfrage

bedeuten würde, als eine verstärkte Erschließung der Peripherie. Bei genauerer Untersuchung der Wirkeffekte lässt sich dies insbesondere auf eine vorteilhaftere Raumstruktur des Stadtkerns und die positiven Dichteeffekte zurückführen, die gemeinsam die dort ebenfalls stärkeren Kannibalisierungseffekte mehr als wettmachen.

Zudem wurden die Stationen des Validierungsdatensatzes mit den 48 Regressionsgleichungen zum relativen Tagesgang der Entleihen und Rückgaben an Werktagen und Wochenenden gescored. Die Abbildung 3 zeigt eine zufällige Auswahl dieser Prognosen (grün) sowie derer tatsächlichen empirischen Verläufe (schwarz). Bei einem Großteil der (disjunkt zu den Trainingsdaten gezogenen) Validierungsstationen ist eine sehr hohe Passgenauigkeit beobachtbar, teilweise verlaufen beide Kurven nahezu deckungsgleich. Dies spricht für eine hohe Prognosekraft des verwendeten Ansatzes. Zudem zeigt sich das Modell deutlich flexibler, als eine bloße Zuordnung zu Balanceclustern in [15] und [16].

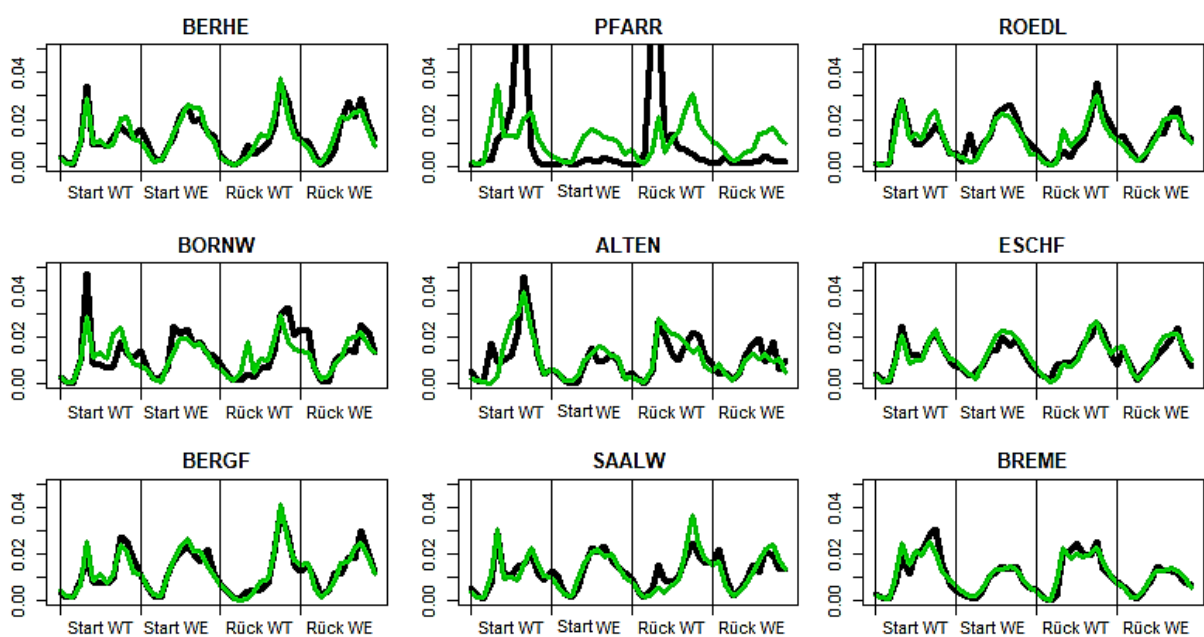


Abbildung 3: Prognostizierte (grün) und empirische (schwarz) relative Nutzungshäufigkeiten im Tagesgang einiger zufällig gezogener Stationen des Validierungssets: Entleihen („Start“) und Rückgaben („Rück“) an Werktagen („WT“) und Wochenenden („WE“).

Vereinzelte sind jedoch auch starke Abweichungen erkennbar. Die Station *PFARR* ist beispielsweise in unmittelbarer Nähe zu einem großen DB-Bürogebäude gelegen, dessen Mitarbeiter den Service unter Vorzugskonditionen nutzen können. Dieser nicht im Modell abgebildete Einfluss unterstreicht (neben der Bedeutung einer umfangreichen Datenbasis) die Unabdingbarkeit einer fachlichen Produktkenntnis, um die Modellergebnisse einzuordnen und validieren zu können.

5 Fazit

In diesem Beitrag wurde ein dreiteiliges Framework vorgestellt, mithilfe dessen potentielle Stationsstandorte für eine Netzerweiterung hinsichtlich der erwarteten absoluten Nutzungshäufigkeit, der im Bestandsnetz zusätzlich induzierten Nachfrage sowie des erwartbaren Tagesgangs evaluiert werden können. Daten unterschiedlichster Quelle und

Struktur fanden dabei Berücksichtigung und wurden hinsichtlich ihres Einflusses auf das Nutzungsverhalten untersucht.

Wichtige Ergebnisse können folgendermaßen zusammengefasst werden:

Es lassen sich drei verschieden wirkende Netzeffekte identifizieren: Je näher die nächstgelegene Station, desto geringer ist die Nachfrage. Je mehr Stationen in attraktiver Zielentfernung und je größer das Netz insgesamt ist, desto höher ist die Nachfrage. Eine Verdichtung des Kernnetzes erhöht vor diesem Hintergrund die erwartete Nachfrage stärker als die Erschließung weiterer peripherer Lagen. Die absoluten Nutzungshäufigkeiten einer Station sind zumindest teilweise durch andere Determinanten bestimmt, als deren relative tageszeitliche Aufteilung. Somit sind getrennte Schätzungen empfehlenswert. Mittels eines Systems von 48 unüberwacht generierten Regressionsgleichungen lassen sich erwartete Nachfragen passgenauer schätzen, als dies über Zuordnungen zu Nutzungsclustern möglich ist. Die Evaluierung potentieller Stationen sollte anhand der im Gesamtnetz zu erwartenden zusätzlichen Umsätze und Kosten erfolgen. Dieser Beitrag liefert ein Framework für eine umfängliche Prognose verschiedener Nachfrageeffekte. Konsequenterweise sind diese mit Kenntnis der internen Kosten- und Erlösstruktur zu monetarisieren, um so betriebswirtschaftlich optimierte Netzplanung betreiben zu können.

6 Literatur

- [1] Wang, X., Lindsey, G., Schoner, J. E. und Harrison, A. (2015): *Modeling bike share station activity: Effects of nearby business and jobs on trips to and from stations*. Journal of Urban Planning and Development, 142(1).
- [2] Sato, H., Miwa, T. und Morikawa, T. (2015): *A study on use and location of community cycle stations*. Research in Transportation Economics, 53:13-19.
- [3] Faghih-Imani, A. und Eluru, N. (2015): *Analysing bicycle-sharing system user destination choice preferences*, Journal of transport geography 44:53-64.
- [4] Zhang, Y., Thomas, T., Bussel, M. und Van Maarseveen, M. (2017): *Exploring the impact of build environment factors on the use of public bikes and bike stations: case study in Zhongshan, China*, Journal of transport geography, 58:59-70.
- [5] Croci, E. und Rossi, D. (2014): *Optimizing the position of bike sharing stations*, Centre for Research on Energy and Environmental Economics and Policy: Working Paper Series.
- [6] El-Assi, W., Mahmoud, M. S. und Habib, K. N. (2015): *Effects of build environment and weather on bike sharing demand*, Transportation, 44(3):589-613.
- [7] Chen, P., Zhou, J. und Sun, F. (2017): *Built environment determinants of bicycle volume: A longitudinal analysis*, Journal of Transport and Land Use, 10(1).
- [8] Garcia-Palomares, J.C., Gutierrez, J. und Latorre, M. (2012): *Optimizing the location of stations in bike-sharing programs*, Applied Geography, 35(1):235-246.

- [9] Frade, I. und Ribeiro, A. (2015): *Bike-sharing stations: A maximal covering location approach*, Transportation Research Part A: Policy and Practice, 82:216-227.
- [10] Xu, Y., Shaw, S.-L., Fang, Z. und Yin, L. (2016): *Estimating potential demand of bicycle trips from mobile phone data – An anchor-point based approach*, ISPRS International Journal of Geo-Information, 5(8):131.
- [11] Liu, J., Li, Q., Qu, M., Chen, W., Yang, J., Xiong, H., Zhong, H. und Fu, Y. (2015): *Station site optimization in bike sharing systems*, IEEE International Conference on Data Mining (ICDM), 883-888.
- [12] Liu, J., Sun, L., Li, Q., Ming, J., Liu, Y. und Xiong, H. (2017): *Functional zone based hierarchical demand prediction for bike system expansion*, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 957-966.
- [13] Deutsche Bahn AG (2017): *Buchungen Call a Bike und Stationen Call a Bike (Stand 05/2017)*, <http://data.deutschebahn.com/dataset/data-call-a-bike>, 13.02.2018.
- [14] National Aeronautics and Space Administration NASA und Ministry of Economy, Trade, and Industry METI (2016): *ASTER Global Digital Elevation Map v2.*, <https://gdex.cr.usgs.gov/gdex/>, 13.02.2018.
- [15] Bürgeramt, Statistik und Wahlen der Stadt Frankfurt am Main (2015): *Materialien zu Stadtbeobachtung – Band 22*, Frankfurt am Main.
- [16] OpenStreetMap-Mitwirkende (2017): *Kartenausschnitt des Stadtgebietes Frankfurt am Main*, <https://planet.openstreetmap.org/>, 14.08.2017.
- [17] Deutscher Wetterdienst (2017): *Klimadaten Deutschland – Stundenwerte (Archiv):* Station Frankfurt/Main, <https://www.dwd.de/DE/leistungen/klimadatendeutschland/klarchivstunden.html?nn=16102>, 13.02.2018.
- [18] Lohse, D., Teichert, H., Dugge, B. und Bachner, G. (1997): *Ermittlung von Verkehrsströmen mit n-linearen Gleichungssystemen unter Betrachtung von Nebenbedingungen einschließlich Parameterschätzung*, Band Heft 5 in Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr, Dresden Technische Universität Dresden.
- [19] Gelman, A. und Hill, J. (2007): *Data Analysis Using Regression and Multi-level/Hierarchical Models*, Analytical Methods for Social Research, Cambridge u.a., Cambridge University Press.
- [20] Bates, D.M. (2010): *Mixed-effects modeling with R*, New York, Springer.