Commuter classification and behavior clustering: Beijing use case

Selene Baez Santamaria
s.baezsantamaria@student.vu.nl

Abstract

Public transportation, centered on subway and bus networks, is an data-rich domain that can benefit from data mining and machine learning techniques. The classification of commuters versus non-commuters/occasional travelers can help government, transport management and operators to better target their policies in order to improve the transportation network in large cities. Furthermore, characterizing commuters by behavior clustering can bring deeper insight into their needs and routines as a whole. This project proposes the usage of ensemble models for classification and clustering of public transport users. For this purpose, transit card data will be used, available from the city of Beijing, China.

Contents

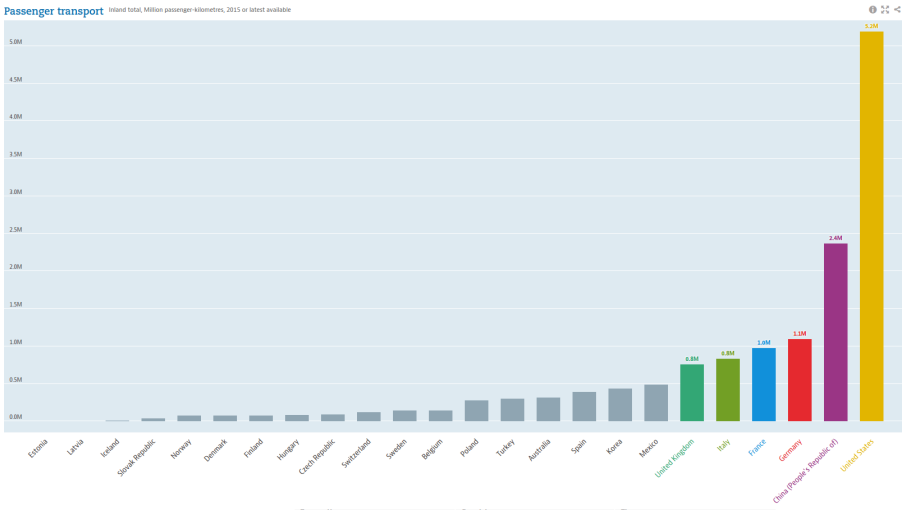
1 Introduction

1.1 Transportation domain

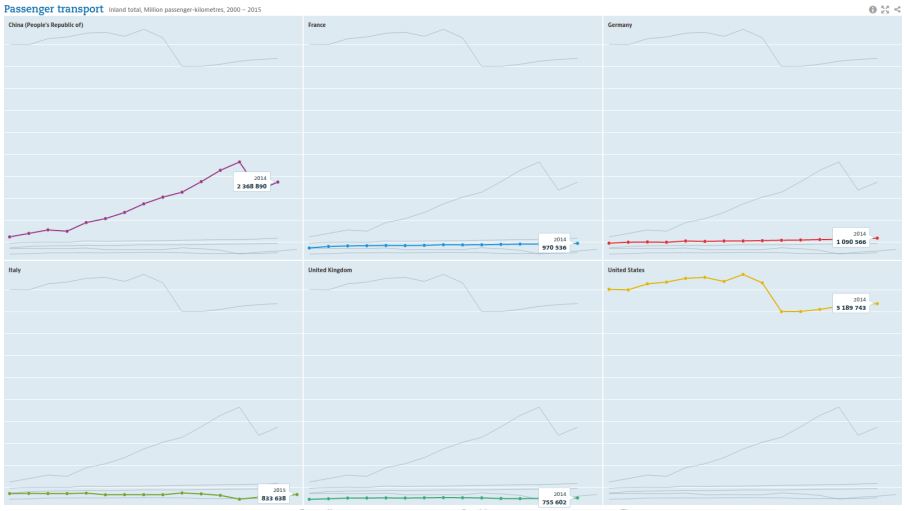
Urban public transportation includes systems that are available for use by anyone in urban areas. Its facilities are commonly composed by buses, subway/metro lines, light rails, tramways, trains, taxis and others. As a network, they provide service for the majority of citizens in urban areas.[20]

Figure 1 shows the passsenger transport usage, as million passengers per kilometer. This represents the transport of a passenger for one kilometer. From the top image we note that United Stated, China, Germany, France, Italy, and United Kingdom contitute the six countries with the most passenger transport, according to their reported data from 2015 or later.[12]

Furthermore, historical data in the bottom image reveals the 15 years behavior for each of the aforementioned countries. Most of the countries show stability, with increase or decrease of less than .10 million passengers for European countries, and .5 million passengers for United States. China, however, shows a trend with steep increase for most of the selected years. In fact, comparing to its less than 1.2 million passengers in 2000, China doubled its public transport usage to 2.4 million passengers in 2015.



(a) Passenger transport data, per country.



(b) Historical data for the top six countries with most passenger transport usage.

Figure 1: OECD countries and their passenger transportation data.

Though it is a more sustainable alternative compared to private car usage, public transport usage has a significant environmental impact, affecting noise and air pollution. Diesel buses, which generally make up a major part of public buses, have large fuel consumption needs and contribute significantly to CO_2 emissions. Even eco-friendly alternatives such as hybrid diesel buses are sensitive to operating conditions, as their fuel consumption may increase by up tp 50% when the on-board air conditioning is on.[23]

Consequently, public transportation directly relates to energetic demand, since its facilities are mostly petroleum or electrical based. In terms of global energy consumption, passenger transportation accounts for about 25% of the total world energy consumption. Furthermore, the tarnsportation sector consupmtion increases at an annual average rate of 1.4.% [7] This may bring further economical implications for countries with high public transportation demand.

1.1.1 Who are the commuters?

A major proportion of public transport users is represented by commuters. These are regular users of public transit, with consistent spatio-temporal patterns in their travels. Driven by a routine, commuters travel back and forth from specific places, commonly represented by their home, work, school, or other similar locations.

As commuters are frequent users of public transit, the conditions of the public network directly influence commuters' personal well being and generally impacts their quality of life. Intuitively, if the commuting experience is unpleasant, daily travel can bring distress to commuters and even repel them from using the public transport at all. Several studies have looked into public transit evaluation from different perspectives, including commuters' needs. The most common aspects of it include: travel time, average speed, delays, accessibility, service coverage, crowded level, facilities quality, and fare rate. Weng et al [21] identified five categories (Convenience, rapid, Reliability and Comfort) that summarize commuters priorities when choosing to travel by public transport.

From both of the above, the large presence of commuters and their known needs, it follows that identifying commuters can help in creating a sustainable public transportation network. Long term planning and policies for improving the overall experience and aid urban areas on a large scale.

Transportation follows swarm behavior. Based on individual travels and routines, on a larger scale travelers exhibit peculiar characteristics. Both levels of understanding are crucial. [1]

1.2 The city of Beijing

The city of Beijing presents a special case of urbanization and rapid industrialization. This is reflected in a sudden population growth of 20% per decade since 1960, with the largest increase of 44% in the last ten years. The latest official census in 2010 reported that the Urban agglomeration of Beijing (including Beijing itself and its adjacent suburban areas) had population of 19,612,368. The UN World Urbanization Prospects estimates the 2017 population at over 22 million inhabitants. [17]

Air pollution in Beijing has become an environmental issue due to its pollutant emissions. [22]

Beijing public transport is composed of buses, subway and bicycles. It is continuously expanding.

Bus: In 2015, there were 876 bus lines with 23,287 buses in operation. The bus network is the most extensive mode of transportation, expanding over 20,186 km. It observes an average daily traffic volume of 10.98 million passengers, with the highest daily volume reaching 13.07 million on one day. [3]

Subway: The Beijing subway has 18 lines with 334 stations, of which 53 are transfer stations. In 2015 it had an operating length of 554 km, with 5,024 vehicles running. [3] Its network is is split by two operators: the state-owned Beijing Mass Transit Railway Operation Corp (operating 15 lines), and the joint Hong Kong venture Beijing MTR Corp (operating 3 lines).

Beijing's subway has an average daily traffic volume of 9.11 million passengers, with a maximum recorded volume of 11.66 million passengers. As such, it is the second busiest metro system in the world, providing 3,410 million annual journeys. Compared to the

[1] expand on this

service provided in 2012, the system observed a 39% increase in usage by 2014. It is also the second longest metro network, surpassed by Shanghai by only 21 km. [19]

Bicycles: Beijing first implemented public bicycle systems in 2012. As of 2015, in total, 67,000 bikes are available for rental with 2,700 pick up/drop off points spread across the city. [3]

1.3 Motivation

Interdisciplinary study between Artificial Intelligence and Metropolitan Transportation. Introduce data mining techniques to a data rich domain.

Many prediction algorithms available. Constant refinement, state of the art must be applied to real life and large impact situations. Domain experts must focus on analyzing insights and using them, not on techniques for curating and making sense out of raw data.

Relevance of project on both areas.

1.3.1 Societal context

A survey conducted in 2009 showed that 80% of the public transport passengers' complaints in Beijing were related to the network being slow and time-consuming, inconvenient to transfer, unpunctual and unreliable. Commuters use the public transport network regularly to perform their routines, and thus need efficient and reliable means of transportation. Flaws and failures in the network reduce the attraction of public transport. [13]

The city of Beijing faces a large imbalance between residential and working areas. [2]. Targeting this group brings the largest benefits to the public.

Government, transport management and operators can gain spatial and temporal insight. This insight can lead to tangible results, policies and counter measures increasing efficiency of network, adjustable travel fares used as incentives to relieve peak hours, urban planning for residential and industrial land use, and others [3]

1.3.2 Scientific context

Usage of machine learning of data mining has been limited. Current broadly used method is surveys to reach travelers on individual level and aggregated measurements for gathering their collective behavior. The analysis is usually done with statistical methodology.

Surveys are costly and based on self-report, which by itself has bias problems. Other problems are small population and non-representative samples.

Aggregated methods miss the interactions between individuals that cause the collective behavior.

Big Data In the last years, smart card systems have become more popular, making it possible to monitor travelers transactions and facilitating fare collection. Several cities have implemented such systems, for example the Octopus card in Hong Kong[5], Oyster card in London [2], OV-chipkaart in The Netherlands [6], and Yikatong card in Beijing [4], to name a few.

Incentive for using Yikatong card since bus is half the price. Over 90% of public transit users are smart card holders. [9]

Technology has reached the data collection point, but has yet to reach the analysis part. Transit cards are capable of recording spatio-temporal information at an individual level over long periods of time. This generates large amounts of historical data that can be mined.

1.4 Thesis organization

This Thesis is organized as follows:

First we do a literature review for previous work on mining transit data and for specific state-of-the art methodologies. Consequently, we establish the scope and objectives of this project. We continue to

[2] reference
[3] reference

describe the methodology thoroughly, including the data and the approach. Following this description, we identify three distinct stages of the project and report their corresponding experimentation. Then, we discuss the findings and gather conclusions. Finally, future work opportunities are explored.

2 Literature review

2.1 Data mining on transit card data

Morency et al. looked at spatio-temporal information in Canadian smart card data. On the one hand, they examine spatial variability by measuring the number of distinct stops a smart card user visits, and the frequency of each stop. On the other hand, they examine temporal variability by clustering the boarding times of each smart card into fours clusters using the k-means algorithm. [11] [4]

Density Based Scanning Algorithm with Noise to classify travelers according to their travel patterns. [10]

Bhaskar et al. did passenger segmentation using only smart card data. They use a two level DBSCAN algorithm for investigating spatial features, such as frequency in Origin-Destination pairs. Separately, they applied DBSCAN to temporal features to determine most frequent boarding times. [1] [5]

Machine learning for commuters identification. SVM with 94% accuracy. [18]

Langlois et al. [8] present an innovative representation for smart card data. Using four weeks worth of data from London Oyster cards, they represent the card information as a time-ordered sequence of inferred activities. Thus, a three dimensional matrix is created where x represents the day in the four week period, y represents the hourly time bin, and z represents the area where the inferred activity took place. The authors perform Principal Component Analysis (PCA) on this matrix in order to engineer the features that characterize the data. 20 bootstrapped samples were taken to analyze the average correlation of the first 13 components, resulting in the selection of the first 8 components as the most informative and stable. The projections of a user sequence onto these components constitute the features to be clustered using k-means. 11 clusters are found and characterized by evaluating demographic variables using odds ratio. Authors further grouped the clusters under "working day", "home bound", "complex activity pattern" and "interrupted pattern" categories.

The latest work on the field corresponds to Ma et Al. [9] The objective of their work is to determine a scoring function for travelers that can correctly identify them as commuters, or non-commuters. In their work, they cluster stops using an improved DBSCAN algorithm. They engineer features for representing the frequency in which travelers follow spatio-temporal patterns. Travelers are then clustered according to these features following the ISODATA algorithm. As an output of the clustering, optimal cutoff levels in the scoring function were determined. As a result, evaluating a traveler does not depend on clustering centroids, but only on calculating the commuting score. This, as expressed by the authors, reduces computing time and treats each traveler independently from the others, which is not true for clustering algorithms.

A common practice, as used by [9], [8], and [11] is to divide the day into -hourly or half-and-hour-time bins.

Volume of data: The volume of data analyzed by previous work ranges from hundreds of smart cards to tens of millions of smart cards, leading to up to hundreds of millions of individual smart card transactions. The details are summarized in Table 1.

Authors	Year of publication	Records	Unique smart cards	Time span
Tu et al. [18]	2016	Unknown	978	one week
Morency et al. [11]	2007	2.2 million	7,118	277 days
Langlois et al. [8]	2016	3 million	33,026	four weeks
Bhaskar et al. [1]	2015	34.8 million	1 million	4 months
Ma et al. [10]	2013	Unknown	3 million	one week
Ortega [14]	2013	65 million	5.7 million	one week
Ma et e. [9]	2017	364 million	18 million	one month

Table 1: Volume of data analyzed by different authors

[4] revise
[5] revise

2.2 Classifying and clustering spatio-temporal data

Clustering algorithms:

- Hierarchical clustering:
  Langlois use it for areas clustering. In order to infer the user-specific activities, all stops or stations visited by each user are clustered using an hierarchical clustering algorithm that considers the distance between stops and the frequency of travel between them. Therefore, different activities are likely to be associated with different areas. [8]
- Partitional clustering:
  Density based [1]
  k-means [11]. Langlois et al. tune the number of clusters k using the DB-index [8].
  k-nearest neighbor [14]

3 Research objective

The objective of this project is to identify and characterize commuters in the city of Beijing by using smart card data. As such, the underlying goal is to find patterns in the spatio-temporal data of public transport travelers, and interpret the results using domain knowledge by experts in the Transportation domain.

3.1 Research questions

The min objective is further broken down into answering the following research questions:

1. How accurately can commuters and non-commuters be identified using an ensemble model? How does this compare to the previous SVM model?
2. What is the minimal set of information needed from smart card data to reach an acceptable accuracy in classification?
3. To what extent is clustering commuters by its behavior informative to transportation specialists?

3.1.1 Definition of terms

A commuter is a public transit user whose smart card data reveals repeatable patterns in time and space.

A trip is a sequence of smart card transactions, including transfers, with an origin and destination. A trip is also represented as a record in the data, as it will be further explained in Section 4.1

A transfer is a change in transportation mode. Transportation modes include Bus, Subway, and Bike.

We make the assumption that smart card IDs and users have a one to one relationship, meaning each user has exactly one card and each card is used by exactly one user. As discussed with domain expert Quian Tu, although some people may own more than one card, this is a minority.

3.2 Scope and structure

This project is divided three main stages:

PART I Classify commuters versus non-commuters by using an ensemble model. We examine the performance of our model and make a comparison with Tu's SVM model [18]. This is explored in Section 5.

PART II Revise features used in previous model in order to identify the most informative features and disregard redundant information. An extensive analysis of spatio-temporal properties will be done, combining transportation domain knowledge and statistical tools. This is explored in Section 6

PART III Commuters will be further clustered according to patterns in their travel behaviors. The clusters will be analyzed and interpreted to find distinctive characteristics that may be deemed as informative for transportation specialists. This is explored in Section 7

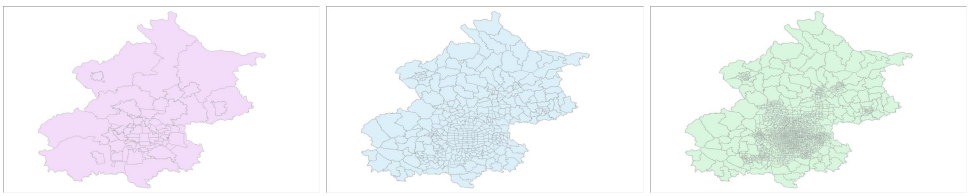
4 Methodology

4.1 The data

Every record int he data represents a trip performed by a specific smart card. As such, it contains the following data fields:

- Data date: Year, month and day that the trip was made
- Card code: card identification number
- Path link: Mode of transportation. B for bus, R for subway, Y for bicycle. Transfers between modes are shown by a dash. [6]
- Travel time: Time spent in vehicles, measured in milliseconds
- Travel distance: Distance traveled, measured in meters as measured by route.
- Transfer number: Number of changes in travel mode during the trip. Regarding transfers between same travel mode, bus to bus, and bicycle to bicycle transfers are counted. Subway to subway transfers are ignored.
- Transfer average time: Time spent in transfer, divided by number of transfers. , Measured in milliseconds
- Transfer total time: Total time spent in transfer, measured in milliseconds
- Start/End time: Time stamp of when the trip started/ended. Date and time with milliseconds precision
- On/Off small traffic area: Integer from 1 to 1911
- On/Off middle traffic area: Integer from 1 to 389
- On/Off big traffic area: Integer from 1 to 60
- On/Off ring road: Integer 1 to 6
- On/Off area: Integer from 1 to 18
- ID: record identification number created by joining the following: hour | time stamp of beginning of trip | card code
- Transfer detail: Mode of transportation, as well as line/route number and stations for boarding and alighting. More detail provided in Section 4.2.2

The traffic zones (small, middle and big areas) are divided by the Beijing Municipal Institute of City Planning and Design (BICP). They are specific in different degrees, as shown in Figure 2. In general, the division principles correspond to the geopolitical environment and administrative planning, for example roads, villages and others. The 6 ring road and 18 areas districts are divided by the Beijing Municipal Government. The division is unique in Beijing. The 18 districts and counties are shown in Figure 3. According to domain expert PhD. Liang Quan, these divisions are sufficiently informative for traffic analysis [16].



(a) Big zones (b) Middle zones (c) Small zones

Figure 2: Traffic zone division

Every day, more than 13 million records are collected, with approximately 5 million corresponding to subway trips, 8 million corresponding to bus trips and 100,000 corresponding to bicycle trips.

[6] Example: B-B is Bus to Bus.

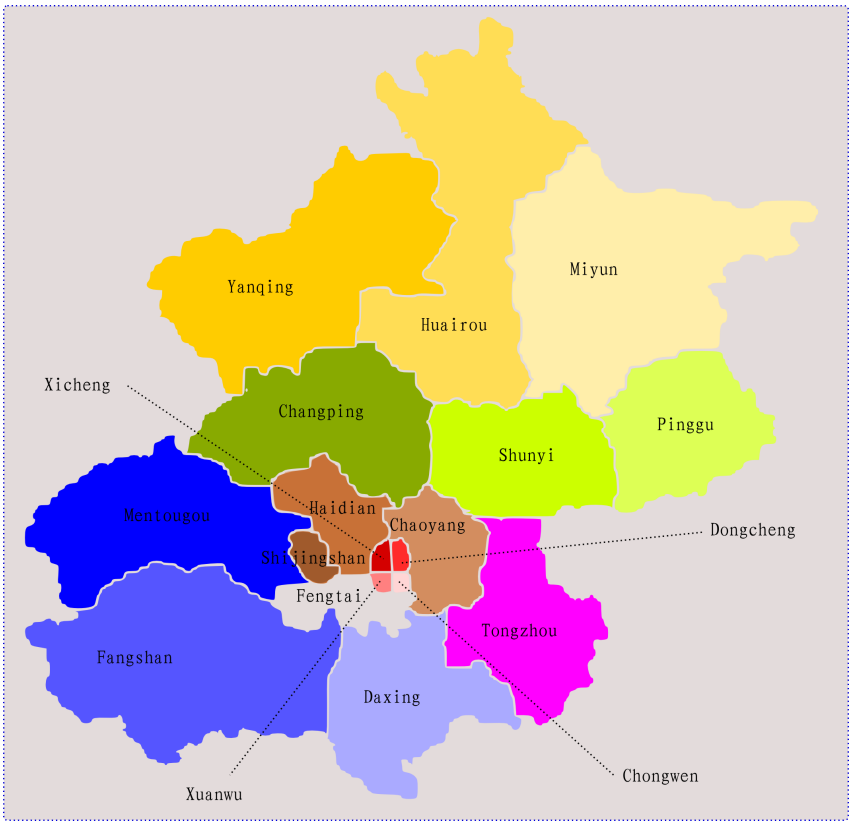


Figure 3: Beijing's Districts and its Counties

4.1.1 Special considerations

The previous description corresponds to the data as delivered by the Beijing Transportation Research Centre. As such, it has undergone some processing from the raw collecting phase. Some special considerations are explained below:

Travel distance by bike: Since bicycles do not have predefined routes, the distance cannot be directly recorded. However, it is inferred by using the travel time and a static average speed for cyclist.

Subway transfer: Transfers between subways lines cannot be tracked since a single check-in gives access to the traveler to all the subway network. In order to infer the transfer detail, the A* algorithm is used to calculate the most likely transfer sequence, given the boarding and alighting stations. Similarly to the bicycle missing information, the transfer time cannot be directly recorded. Using a static average walking speed and the known distance in transfer stations, the transfer time is calculated.

Transfer information: The path link and transfer number fields are extracted from the transfer detail field. Similarly, the transfer average time is calculated from the transfer total time and transfer number fields.

4.1.2 Labeled and unlabeled data

Commuter classification: Classification is a supervised learning task, where every training data sample requires an associated label determining its true class. In case of commuter classification, this translates to having smart card codes associated with either a "commuter" or "non-commuter" label. Such data is expensive and limited since it can only be obtained by asking the users directly if they are commuters or not. Thus, in general, annotated data is not available, and labeling new records falls beyond the scope of this project.

As a solution for the above, we take advantage of the dataset used by Tu[18]. This dataset corresponds to trip records performed during a week in January 2015, and it contains labels for 978 smart cards, collected and validated via surveys. The original dataset distribution is composed by:

- 6439 records of 481 commuters
- 1628 records of 497 non-commuters

For this project, the Beijing Transportation Research Centre has provided us with one month worth of data, corresponding to January 2015. In order to construct an extended labeled dataset, we take these 978 labeled smart card IDs and search for their corresponding records in the one month sample. This dataset is used for Part I (Section 5) and Part II (Section 6) of this project.

Commuter clustering: For this project, 100, 000 records are sampled every day for a month. In this case, the samples correspond to November 2014, which does not overlap with holidays and has a relatively stable weather thus diminishing the variance between bicycle and bus/subway traveler preferences.

4.2 Data preprocessing

4.2.1 Cleaning

As first step for preprocessing the data, we perform a cleaning where we eliminate records that are faulty, for example:

1. Eliminate records with missing data: 10.9% records eliminated
2. Eliminate records with travel time <= 0: <0.01% records eliminated
3. Eliminate records with travel distance <= 0: 10.5% records eliminated
4. Eliminate records linked to users with insufficient trips: 25.8% records eliminated when requiring at least 2 trips.

This leaves 52.6% records available for usage.

Insufficient trips regulated by a user input [7]

4.2.2 Trip parsing

We parse the trip details using a combination of two techniques: regular expressions and tokenization.

Regular expression In order to obtain the elements of the trip

$$
\begin{aligned}
BIKE &= (bike.STATION - STATION) \\
SUBWAY &= (subway.LINE_{NAME} : STATION - LINE_{NAME} : STATION) \\
BUS &= (bus.ROUTE_{NAME} :)
\end{aligned}
$$

$$
GENERAL = (MODE.[LINE_{NAME} :]?STATION - [LINE_{NAME} :]?STATION-)
$$

$$
\begin{aligned}
LINE_{NAME} &= 5numberlineorNAMEline \\
ROUTE_{NAME} &= 944ornight32(DIRECTION - DIRECTION)
\end{aligned}
$$

Tokenization Dictionary [8]

Example in Chinese -> English -> clean trip

[7] plot percentage of records and min records needed. Tune parameter

[8] encode chinese in latex

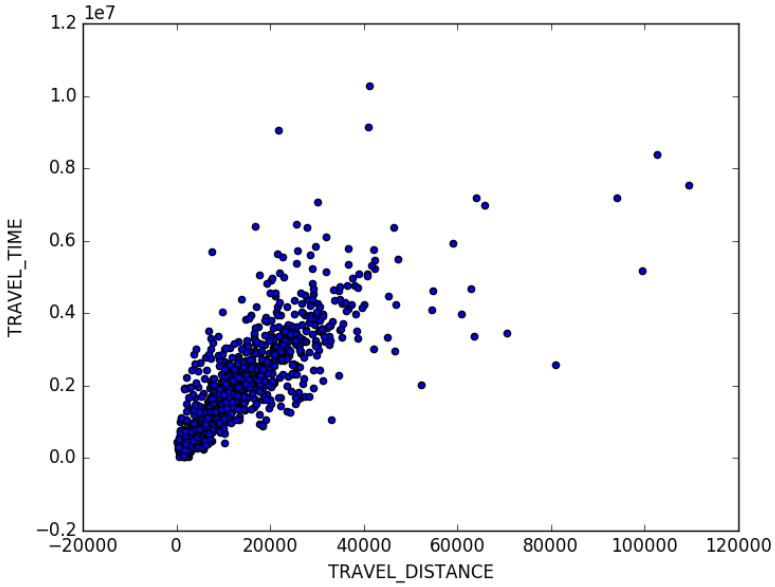


Figure 4: Travel distance vs travel time.1000 record sample.

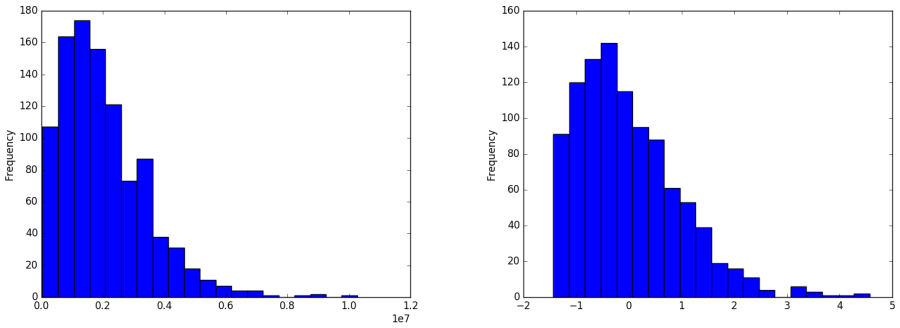
4.2.3 Data patching

We note that num of transfers and path link are faulty, According to expert PhD. Tu Qiang, this must be recalculated [15]. [9]

4.2.4 Transformation and Standardization

Whitening vs standardization: whitening eliminates correlations between the data. In our case those are relevant and we ant to keep them. [10]Figure 4 shows a clear correlation between travel distance and travel time.

Travel time and distance where standardized by subtracting the mean and forcing a unit standard deviation. [11] [12]

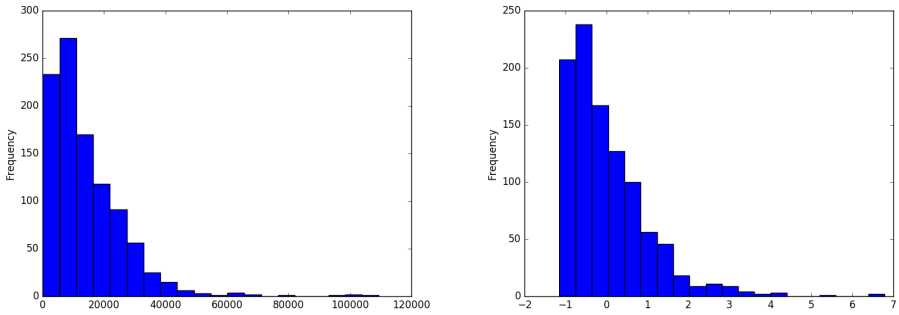


(a) Original

(b) Standarized

Figure 5: Time distribution before and after preprocessing. 1000 records sample.

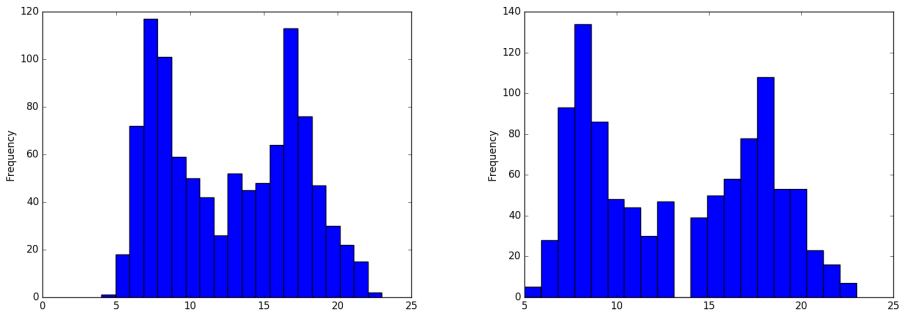
[9] recalculate num of transfers, and path link.
[10] Correlation between time and distance to be preserved?
[11] double check this
[12] replace images once run with all data



(a) Original (b) Standarized

Figure 6: Distance distribution before and after preprocessing. 1000 records sample.

Hourly time bins is standard practice in the field [8] [9],



(a) Start hour (b) End hour

Figure 7: Distribution of start/end hours for trips.1000 records sample.

Figure 7 shows clear morning and evening peak hours.

Furthermore, we create a marker to determine if the record was conducted over a weekday or not. [13]

4.3 Data mining techniques

The work for this project is coded in Python, using libraries for data and numerical manipulation (*Pandas*, *numpy*), regular expressions (*re*), machine learning (*sklearn*), neural networks(*Theano*).

4.3.1 Feature engineering

The temporal factors to be explored are represented by the start/end times, as well travel/transfer time.

The spatial factors to be explored are represented by On/Off areas. [14].

[15]

Feature engineering by deep learning

[13] TODO: implement in code

[14] and route lines?

[15] hierarchical features? data structure for this?

4.3.2 Ensemble models

Ensemble models are chosen because of its robustness and modularity. Starting from two simple classifiers, assembled via bagging, the model can grow larger or more complex as needed and it may be extended beyond the scope of this Thesis Project.

[16]

4.4 Correlation analysis

chi-test

[16] ensemble models for supervised learning, consensus clustering for unsupervised

5 Commuters identification

5.1 Hypothesis

As suggested by Tu [18] results, the data is almost linearly separable thus simple classifiers such as decision trees may suffice.

5.2 Model

A first instance of the model will use all available variables in the data as used by Tu [18] for a fair model comparison.

With data used by Tu[18], random forest achieves %99.96 accuracy. Probably overfit.

5.3 Experiments

5.4 Results

Accuracy

Confusion matrix

6 Variable evaluation

6.1 Hypothesis

One of the main focuses of the second phase of this thesis is to determine the appropriate level of detail in the area to be taken into account.

Small, middle and bug area overlap. Middle and small divisions have more precision but maybe are not needed. On the other hand, big area divisions might not capture the changes for people who live and work/study in the same district.

6.2 Qualitative

Exploration: Experts opinion

6.2.1 Interview

[17] We interview Liang Quan as an Transportation domain expert.

- To what extent do people live and work on the same area?
- what level of detail do you think is appropriate?

6.3 Quantitative

Analysis: Correlation

[17] In appendix?

7 Commuters clustering

7.1 Feature engineering

7.1.1 Convolutional filters

Image like structure. Inspired by [8]

x, y dimension are temporal -> day time z dimension is spatial -> areas

Matrix of 30 x 24 x 15 (for all features or 10 for only spatial)

7.2 Neural networks

Reduce dimensionality from 10,800 to 200
[18]

7.3 Model

consensus clustering

7.4 Experiments

7.5 Results

7.6 Expert judgment

[18] visualize features

8 Conclusion

9 Future work

References

[1] Ashish Bhaskar, Edward Chung, et al. Passenger segmentation using smart card data. *IEEE Transactions on intelligent transportation systems*, 16(3):1537–1548, 2015.

[2] Philip T Blythe. Improving public transport ticketing through smart cards. In *Proceedings of the Institution of Civil Engineers, Municipal Engineer*, volume 157, pages 47–54. Citeseer, 2004.

[3] Beijing Transportation Research Center. 2016 annual report on traffic development in beijing. `http://www.bjtrc.org.cn/InfoCenter/NewsAttach/2016%E5%B9%B4%E5%8C%97%E4%BA%AC%E4%BA%A4%E9%80%9A%E5%8F%91%E5%B1%95%E5%B9%B4%E6%8A%A5_20161202124122244.pdf`, 2016. Acessed on 21 April, 2017.

[4] Mo Lim Chan. *Tactical implementation model for the smart card payment system for metro operation*. PhD thesis, City University of Hong Kong, 2010.

[5] Patrick YK Chau and Simpson Poon. Octopus: an e-cash payment system success story. *Communications of the ACM*, 46(9):129–133, 2003.

[6] Gerhard de Koning Gans. *Analysis of the MIFARE Classic used in the OV-chipkaart project*. PhD thesis, Master's thesis, Radboud University Nijmegen, 2008.

[7] US EIA. Energy information administration (2016), international energy outlook 2016, with projections to 2040. Technical report, DOE/EIA-0484, 2016.

[8] Gabriel Goulet Langlois, Haris N Koutsopoulos, and Jinhua Zhao. Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, 64:1–16, 2016.

[9] Xiaolei Ma, Congcong Liu, Huimin Wen, Yunpeng Wang, and Yao-Jan Wu. Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*, 58:135–145, 2017.

[10] Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36:1–12, 2013.

[11] Catherine Morency, Martin Trepanier, and Bruno Agard. Measuring transit use variability with smart-card data. *Transport Policy*, 14(3):193–203, 2007.

[12] OECD. Passenger transport (indicator). `10.1787/463da4d1-en`, 2017. Acessed on 10 April, 2017.

[13] Beijing Municipal Commitee of Communications and Beijing Transportation Reserch Center. Research on beijing's public transportation commuting transit network. `https://www.esmap.org/sites/esmap.org/files/10282009102930_Beijing_Transport_finalReport.pdf`, 2009. Acessed on 21 April, 2017.

[14] Meisy Andrea Ortega-Tong. *Classification of London's public transport users using smart card data*. PhD thesis, Massachusetts Institute of Technology, 2013.

[15] Tu Qiang. personal communication.

[16] Liang Quan. personal communication.

[17] World Population Review. Beijing population. `http://worldpopulationreview.com/world-cities/beijing-population/`, 2016. Acessed on 21 April, 2017.

[18] Qiang Tu, Jian-cheng Weng, Rong-Liang Yuan, and Peng-fei Lin. Impact analysis of public transport fare adjustment. *Traffic Engineering & Control*, 57(2), 2016.

[19] UITP. World metro figures, statistics brief. `http://www.uitp.org/sites/default/files/cck-focus-papers-files/UITP-Statistic%20Brief-Metro-A4-WEB_0.pdf`, 2015. Acessed on 21 April, 2017.

[20] Vukan R Vuchic. Urban public transportation systems and technology. 1900.

[21] Jiancheng Weng, Yueyue Wang, Jianling Huang, and Ledian Zhang. Bus operation monitoring oriented public transit travel index system and calculation models. *Advances in Mechanical Engineering*, 2013.

[22] Hefeng Zhang, Shuxiao Wang, Jiming Hao, Xinming Wang, Shulan Wang, Fahe Chai, and Mei Li. Air pollution and control action in beijing. *Journal of Cleaner Production*, 112:1519–1527, 2016.

[23] Shaojun Zhang, Ye Wu, Huan Liu, Ruikun Huang, Liuhanzi Yang, Zhenhua Li, Lixin Fu, and Jiming Hao. Real-world fuel consumption and co 2 emissions of urban public buses in beijing. *Applied Energy*, 113:1645–1655, 2014.