Article

Relationships between individual characteristics and real walking durations from departures to transit stations

Qi Chen 1, Siting Chen 2, Go Higuchi 3 and Shichen Zhao 4, \*

1 Graduate School of Human-Environment Studies, Kyushu University, Japan; chenqi.clever@gmail.com

2 Graduate School of Human-Environment Studies, Kyushu University, Japan; miche.siz92@gmail.com

3 Graduate School of Human-Environment Studies, Kyushu University, Japan; goooooooooo217@gmail.com

4 Faculty of Human-Environment Studies, Kyushu University, Japan; zhao@arch.kyushu-u.ac.jp

**\*** Correspondence: zhao@arch.kyushu-u.ac.jp

Received: date; Accepted: date; Published: date

**Abstract:** Walking duration is one of the determinants in choosing to use rail transit. This study aims to examine the prevalence of walking durations with respect to different individual characteristics among the residents in Fukuoka of Japan. Walking duration, however, represents the distance between departure and station, it is directly determined by the location of departure but not the individual characteristics. To examine the prevalence of walking duration, this study treats the behavior that walking to rail transit as a travel choice result. For a given walking duration, people with different individual characteristics are supposed to have different willingness in deciding whether to use rail transit. Based on this supposition, the research question can be converted into a binary classification problem, of which the classification is the travel choice, the features are the corresponding individual characteristics of passengers. This binary classification problem is estimated using the random decision forests at three walking durations of 5, 8, and 13 minutes. The result shows that trip purpose plays the most important role in deciding to walk to rail transits. The other features of peak hour, age, occupation also show significance in affecting the willingness of walking to rail transits. This result can be applied to estimate the willingness of walking to rail transit at any given walking duration. This willingness is an important reflection of potential passenger around stations, which can be also used for estimating the catchment area of transit stations.

**Keywords:** walking duration; rail transit station; catchment area; individual characteristics

1. Introduction

At present, the 800m (half-mile) walking distance has been widely accepted as a principal reference of the catchment area for the planning of Transit-Oriented Development (TOD) [1–4]. Planners and researchers also use rail transit catchment areas to make prediction of ridership. In practice, this 800m walking distance is loosely obtained from the sampling survey by asking how far people are willing to walk to rail transit stations, and the same reasoning has been used to justify other rail transit catchment areas and even in different cities and countries with various walking distance preferences. As to passengers, the acceptable walking distance/duration should not be only associated with the features of the built environment such as the scale of stations and the functions of the buildings around stations, but also the features of passengers' individual characteristics, such as occupations and trip purposes.

Rail transit provides a cheaper and environment-friendly way of transportation to passengers. Like any other service and commodities, it needs to face the market and should be transacted at the cost acceptable to consumers. Accordingly, the distribution of transaction cost can reflect the consumption-ability of consumers in the market. Obviously, it is important for rail transit operators to know how much the cost that consumers are willing to pay, based on which thereby making rail transit more attractive. But a little different from general service and commodities, for passengers, the cost refers to the convenience rather than the fare, because the cheap enough fare is already not the important element for passengers to decide whether using rail transit while the walking duration to stations become the important determinant.

As to the case of rail transit, before a potential passenger makes the decision of walking to a station, the expected walking duration, which is the "price" for this potential passenger, is just the reflection of the distance between departures and stations; while if this expected walking duration can be accepted by this potential passenger, this trip will happen and be surveyed. Based on the analogy given before, it's easy to think of that people with different individual characteristics generally should have different willingness towards walking duration to stations [5], and the surveyed walking durations can be viewed as the transaction price which have been accepted by passengers. It follows that the willingness of using rail transit will increase if the expected walking duration is less than that of acceptable range, otherwise, the willingness will decrease [6]. Reflected in the survey data, the number of records with shorter walking duration will be more than that with longer walking duration.

As interpreted above, although the walking duration is just the reflection of distance between departures and stations, it should have relevance with individual characteristics [7–12]. With the aim of explain walking duration to transit stations, this study uses Fukuoka of Japan as the study case to investigate walking durations to transit stations with respect to different individual characteristics. The dataset includes passengers' individual characteristics and trip chain information, for example gender, age, occupation, departures, destinations and travel time. The content is organized as follow: section two proposes the specific research questions based on the reviews of previous research, and proposes some basic assumptions for this study; section three presents the description and disposal of data; section four interprets the process of converting the research problem and the model construction; section five is the estimation results; section six discusses the results and draws conclusions.

2. Literature Review

This section reviews the literature on the issue of walking duration/distance to rail transit stations, some problems still not well addressed are summarized. The review is arranged into two parts, they are, the understanding and disposal of walking duration/distance, and the effect of influencing factors.

2.1. Walking duration/distance

For the walking duration/distance, which is the research object in this study, there is always a difficult point in obtaining the accuracy walking duration/distance by questionnaire because of the discrepancy between perceived values and objective values. Also, the observed walking duration/distance is not the reflection of how long people are willing to spend on walking to rail transit stations, but only the walking duration/distance between departures and stations. For this problem, some studies chose a different perspective trying to explain the walking duration/distance by introducing a threshold of walking duration [7,13]. They examined the differences in the features of explanatory variables in terms of the specific threshold of walking duration. Indeed, investigating the prevalence of walking duration at specific threshold can decrease the discrepancy between observation and reality in some extent, but this disposal also brought some new problems in, for example, it may lead to a great loss of information in the raw data, moreover deciding the threshold of walking duration/distance is also difficult. Another point for this issue is whether to choose walking distance or walking duration as the threshold. To date, there are a lot of studies working on the relationship between walking distance and passengers' individual characteristics, some of them argued that an individual has a limited amount of time spending on traveling during a day, people tend to accept further walking distance as the speed of travel increases [10,14]. With this reasoning it is easy to think of that passengers with the same individual characteristics may have the similar willingness of walking duration, but they generally have different willingness of walking distance due to different travel speed.

2.2. Individual factors influencing walking duration

Among the influencing factors for walking durations, passengers' individual characteristics are generally thought to be the key that can affect walking distance [7–9,11,12,15], whereas, there are few studies having clearly verified the relationship between individual characteristics and walking distance/duration, even there is a study suggesting the walking distance should not be viewed as a function of socio-demographic characteristics [12]. Several studies have confirmed the role of travel purposes in determining walking distance, the commute trip showed particularity from the other purposes, people with the purpose of commute tend to walk a longer distance to rail transit stations [10]. However, the definite relationship between trip purposes and walking distance is still unclear. The situation is the same with other categories of factors, such as the factors of transportation environment, land use, and willingness of passengers [9,12,15]. The only thing that has been confirmed to date is that the walking distance/duration can be influenced by some specific kinds of factors, such as socio-demographic characteristics, trip purposes, and built-environment, but the problem is how and to what degree the walking distance/duration can be influenced.

2.3. Summary and basic assumptions

It may be because of the problems in the disposal of walking distance/duration or the selection of influencing factors, most of the existing studies did not find the significant relevance between the walking distance/duration and the influencing factors. The studies working on the qualitative description for the distribution of walking distance accounted for the majority although some of the existing studies attempted regression model on this issue, the expected results were not obtained. To avoid the problems mentioned in the review, this study uses the thresholds of walking duration as the research object; for any given threshold of walking duration, the respondent who gives the answer longer than the given threshold can be viewed as this respondent can accept this threshold of walking duration; thus, the surveyed walking duration can be converted into the reflection of passengers' willingness for using rail transit. For the analytical method, since various types of regression model has been verified unsuitable for this issue, instead of finding the linear relevance between dependent and independent variables, this study introduces the approach of machine learning into this issue, and try to explain the relevance between walking durations and individual characteristics from the view of probability.

Based on the statements outlined above, several assumptions are necessary for this study.

* : The distribution of departures and destinations of people with different individual characteristics in Fukuoka, Japan, is random in space.
* : The maximum acceptable walking duration of people with same individual characteristics should subject to the normal distribution.
* : The respondents are viewed as they can accept the walking duration that they answered.

Based on and , when examining the threshold of walking duration , set the proportion of group in the whole surveyed sample as , the proportion of people whose walking durations are under the given threshold of is marked as , the proportion of people whose walking durations are over the given threshold of minutes is marked as . If individual characteristics have no significant correlation with maximum acceptable walking durations, there should be no significant differences among , , and ; otherwise, the three proportion should show significant differences, and the differences will show regularities at different threshold .

Based on , if someone gives the answer minutes, it means this respondent can accept the walking duration of minutes and any walking duration that less than minutes. Indeed, this respondent perhaps can accept the walking duration over he answered, but this study just concerns about if he can accept the given threshold of the walking duration other than how long he can accept.

3. Data

3.1. Study case

The study case of Fukuoka is the sixth largest city in Japan which has the population of more than 1.5 million. The dataset is the Northern Kyushu Area Person Trip, which is conducted about every 12 years. The latest available data is from the 4th survey surveyed by the year of 2005, and the 5th survey is already in preparation from September 2017. Figure 1 shows the research area and the distribution of rail transit stations. By the year of 2005 (the 4th Northern Kyushu Area Person Trip Survey was conducted), there are more than 70 stations located within the city area of Fukuoka, of which the number of JR Kyushu station is 27, Fukuoka Subway station is 35, and West Japan Railway station is 16. Now some new rail transit lines and stations are still under planning and construction. According to the “Survey on the current state of public transportation” published by the Ministry of Land, Infrastructure, Transport and Tourism of Japan, the rail transit system in Fukuoka carries a daily average of more than 0.4 million passengers by 2015, accounting for more than 20% in total motorized travel. Although the rail transit system of Fukuoka is still not a large-scale one at now, it has been playing a crucial role in people's daily travel.

|  |
| --- |
|  |
| **Figure 1.** Distribution of Transit Stations |

|  |  |
| --- | --- |
| **Table 1.** Available Data Contents | |
| **Category** | **Feature** |
| Trip chaining behavior | departure location  departure time  arrival time  transport modes  time spent for each mode  location of bus stop or rail transit station |
| Socio-demographic attributes | age  sex  occupation  trip purpose  vehicle/license holding  address |
|  | |

3.2. Data description

The main purpose of person trip survey is to know the travel trends, thus making a better living environment and providing support for traffic planning. The original data covered the range of all the main cities in Northern Kyushu Area, which has more than 483,000 records of trip chaining behavior. The available data in this study mainly includes trip chaining behavior and socio-demographic characteristics, as shown in the Table 1.

To analyze the walking duration between departures and rail transit stations in Fukuoka, the first step is to extract the valid records of rail transit trip within the city area of Fukuoka from over 480,000 records in the dataset. The procedure of extracting the valid data is divided into 3 steps. Firstly, extracting all the person trip data that surveyed within the city area of Fukuoka; secondly, selecting the trip chaining behavior which contains the rail transit mode; thirdly, filtering the invalid data that with null value or abnormal value. The procedure of data cleaning is shown in Figure 2, at last the valid dataset is reduced to a size of 4254 trips.

|  |
| --- |
|  |
| **Figure 2.** The Age Distribution of Passengers |

|  |
| --- |
|  |
| **Figure 2.** Process of Data Cleaning |

Figure 3 shows the age distribution for walking trips to rail transit stations based on the finally valid dataset. The passengers aged from 25 to 55 account for the majority of the whole passengers, while schoolchildren aged under 15 rarely take rail transit. The distribution graph does not show significant peak values at any specific age group. Figure 4 shows the distribution of real walking durations to stations, it has a mean value of 8.32, and the standard deviation is 2. Notably, there are several peak values at the time of 5 multiples. It is speculated that the peak values may be caused by deviation occurred in the investigation. Since people's feeling about the specific time or number is inaccurate, they are inclined to reply a loose answer when they are asked some questions about the details of walking duration. This inclination will count some of the real walking duration that is near to 5 multiples as the 5 multiples and finally expressed in the result of investigation. Despite the bias between survey and reality, as the second assumption proposed before, passengers are viewed that they can accept the walking duration what they answered, therefore, the peak values are thought available in this study.

|  |
| --- |
|  |
| **Figure 4.** Distribution of Walking Duration |

3.3. Data preprocessing

The procedure of feature selection includes two parts, the first is the disposal of features with low variance, and the second is the univariate feature selection. The first one is used for filtering the features accounting for very little in the total samples, because a too small quantity in the sample cannot present significant influence on the result. In this study, the multi-categorical features in the original dataset are reclassified into larger groups, thus dealing with the features with low variance. The second step is to select the valid features that indeed have relevance with the independent variable, this step is based on statistical tests. In this study, the Analysis of Variance (ANOVA) is used for estimating the importance of each feature.

The feature of trip purpose has 15 subcategories in the original dataset, this study reclassifies them into 5 categories including commuting to work, commuting to school, official business, private purpose (such as shopping, entertainment), and going home. The feature of occupation is reclassified from 14 subcategories into 5 categories as well, they are service, technology, administration, student, and the other. Table 2 reports some of the statistical description for each feature. Overall, the average walking duration to rail transit stations is 8.32 minutes; more than 75% of the passengers walk less than 10 minutes; most passengers walk to stations costing 5-10 minutes. In detail, there are some significant differences in the statistical description for each feature, the differences are summarized as follows.

1. The walking duration in peak hour is longer than off-peak hour.
2. Young and old people are inclined to spend less time on walking to rail transit stations.
3. Passengers with the trip purposes of official business and going home tend to take shorter time in walking to rail transit stations.
4. Passengers with the trip purpose of commuting to work are willing to accept a longer walking duration than that with other purposes significantly.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.** Statistical Description of Walking Duration for Each Feature | | | | | | | | | | |
| **Features** | **Categories** | **Count** | **Prop** | **Mean** | **Std** | **10th** | **25th** | **50th** | **75th** | **90th** |
|  | total | 4254 | - | 8.32 | 4.63 | 3 | 5 | 8 | 10 | 15 |
| Sex | male  female | 2257  1996 | 53.10%  46.90% | 8.41  8.22 | 4.63  4.63 | 3  3 | 5  5 | 8  7 | 10  10 | 15  15 |
| Peak hour | peak  off peak | 2976  1277 | 70.00%  30.00% | 8.51  7.88 | 4.66  4.52 | 3  3 | 5  5 | 8  7 | 10  10 | 15  15 |
| Age | 5-24  25-44  45-64  65- | 543  1992  1407  311 | 12.80%  46.80%  33.10%  7.30% | 8.18  8.36  8.43  7.78 | 4.70  4.42  4.83  4.81 | 3  3  3  3 | 5  5  5  5 | 7  8  8  7 | 10  10  10  10 | 15  15  15  15 |
| Occupation | service  tech  office  student  null | 1256  721  1076  325  875 | 29.50%  17.00%  25.30%  7.60%  20.60% | 8.32  8.58  8.44  8.12  8.03 | 4.57  4.89  4.48  4.73  4.62 | 3  3  4  3  3 | 5  5  5  5  5 | 8  8  8  7  7 | 10  10  10  10  10 | 15  15  15  15  15 |
| Purpose | going to work  going to school  business  private purpose  going home | 2697  287  153  789  327 | 63.40%  6.70%  3.60%  18.60%  7.70% | 8.69  8.19  7.10  7.82  7.17 | 4.51  4.60  5.05  4.78  4.67 | 4  3  2  3  2 | 5  5  3  5  5 | 9  8  5  7  5 | 10  10  10  10  10 | 15  15  15  15  15 |
|  | | | | | | | | | | |

4. Methods

In this study, the variable of walking duration to the rail transit station is a continuous one, while the variables of individual characteristics are multi-categorical, for which the multiple linear regression is considered not applicable. Although there were some studies using the multiple linear regression to estimate the walking duration, the result showed no clear relationship [12]. Based on the assumptions proposed before, when giving a specific threshold of walking duration, passengers who walk longer or less than this given threshold should have different features in individual characteristics. The question can be converted into a binary classification problem that people who walked longer than the given duration is labeled as "actually accepted this given value", while people who walked less than the given value is labeled as "unknown". Passengers are thought to have a higher probability of choosing rail transit when given a threshold less than that they have been accepted. The abstract model describing this relevance is expressed as follow, of which Equation (1) defines the choice result, and Equation (2) gives the probability of at given vector of *.*

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

|  |  |
| --- | --- |
|  | is the walking duration answered by passenger (in minutes). |
|  | is the threshold of walking duration that to be examined (in minutes). |
|  | is a binary variable. According to , means passenger can accept the threshold of ; while indicates that it is uncertain whether passenger can accept this threshold. |
|  | is the vector whose component is the individual characteristics of passengers. |
|  | |

Since the question has been converted into a binary classification problem, for which the models of decision tree, Bayesian, support vector machine (SVM), logistic regression, and neural network are widely adopted. In this study, limited by the volume of sample, also the unknown internal relationship among features, the decision tree model should be a good choice because of the good generalization for different forms of data. Furthermore, to avoid the structure of tree from being too complicated, and to improve the robustness of model, this study will use an improved model of decision tree, the random forest model, to train and test the sample. Random forests model is an extension and improvement for the decision tree model, it is operated by constructing a multitude of decision trees and randomly selecting the features at training time [16,17]. The structure is presented in Figure 5. For the random forest model, the prediction is not obtained from the only one decision tree, but the multitude of decision trees constructed by random selection of features and samples. In addition, to improve the efficiency and accuracy of the model, and reduce the number of invalid branches in the model, it is also necessary to filter the invalid features before estimating the model.

|  |
| --- |
|  |
| **Figure 5.** Prediction Process in Random Forests Model |

Briefly, the general process of this study can be summarized as follows. Firstly, data is extracted on the Fukuoka city-wide, including 4254 samples, and 6 main categories of features. Secondly, the features with importance in explaining the willingness of walking duration are selected by using analysis of variance (ANOVA). Thirdly, random forest model for each threshold of walking duration is estimated, the result obtained from the random forest model is the probability of walking longer than the given threshold of walking duration for each passenger in terms of individual characteristics. Finally, the accuracy of this result is evaluated using the method of simple moving average by descending order of the predicted probabilities.

5. Results

5.1. Valid features

According to the achievements from previous studies, the walking distance to rail transit stations ranged generally from 400 meters to 1000 meters in terms of different city types, travel preference, also the needs of research purpose [9,18–22]. If converting this walking distance into walking duration by using the walking speed of 4.8 km/h, the walking duration would range from 5 minutes to 13 minutes [23]. In this study, more than 40% of the samples walk less than 5 minutes, and the walking duration within 13 minutes covers about 85% of the total. It means that the 5 minutes walking duration can be accepted by most passengers, while a walking duration longer than 13 minutes is not a general acceptable one. In addition, as the average walking duration in this study is 8.32 minutes, the three representative thresholds of 5, 8, 13 minutes are picked as the typical threshold for estimating the relationship between individual characteristics and walking durations. The features with the p-value in ANOVA less than 0.05, which means this feature relevant with the dependent variable at the confidence level of 95%, are picked out and listed in Table 3.

From the result in Table 3, travel purpose is the most important factor in determining the walking durations at all the 3 thresholds. The feature of peak hour is also significant in explaining the walking duration. People tend to walk a longer time to stations at peak hours, while people with private purpose or on the way going home are not willing to choose a station far away. For the details of each threshold, at the threshold of 5 minutes, the features of trip purposes and peak hour play the most important role in determining the willingness of walking duration, where the unemployed people and the elderly people tend to walk less than 5 minutes to stations. As to the threshold of 8 minutes, the importance of age and gender raised in some extent, where people whose age is between 25 and 44 are unwilling to walk more than 8 minutes. Moreover, the middle-aged people tend to walk longer than other groups, where people aged from 45 to 64 can accept walking more than 13 minutes to stations more easily.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3.** Valid Features and the Effect at Each Threshold | | | | | |
| **5 min Threshold** | | **8 min Threshold** | | **13 min Threshold** | |
| Feature | Effect 1 | Feature | Effect | Feature | Effect |
| age over 65 | L | female | M | age 45-64 | M |
| peak hour | M | age 25-44 | M | peak hour | M |
| O\_null 2 | L | peak hour | M | P\_going to work | M |
| P\_going to work 3 | M | P\_going to work | M | P\_private purpose | L |
| P\_official business | L | P\_official business | L |  |  |
| P\_private purpose | L | P\_private purpose | L |  |  |
| P\_going home | L | P\_going home | L |  |  |
| 1 L means the individual with this feature tend to walk longer than the given threshold of walking duration, while M is the opposite meaning. | | | | | |
| 2 O is abbreviated from occupation | | | | | |
| 3 P is abbreviated from purpose | | | | | |
|  | | | | | |

This result of feature selection is also consistent with common sense and partly confirmed by the previous research. Most of the surveyed walking durations distribute around the average walking duration of 8 minutes, it can be considered that walking duration between 5 and 13 minutes is sensitive to individual characteristics. The walking duration more than 13 minutes is not accepted by most people even if they have different individual characteristics, and the threshold less than 5 minutes is also not sensitive to individual characteristics since the 5 minutes threshold is generally accepted by most passengers.

5.2. Model estimation

The valid features (Table 3) are used in the random forest model to estimate the probability that walking longer than the given threshold of walking duration. For estimating the model, the dataset is divided into two parts, 50% of the sample are used for fitting the model thus obtaining the coefficients, the rest 50% are used for testing the ability of prediction. The dependent variable at the given threshold of walking duration to rail transit stations is calculated by Equation (1) based on the mean prediction.

The accuracy of results is evaluated by using the method of simple moving average. Figure 6 is the trend line of probabilities that walking less than the given threshold. The trend line of the surveyed values based on the test set is calculated by the mean probability of a group people who have close predicted values. The trend line is drawn by a descending order of predicted values. From the comparison of predicted values and surveyed values, it can be known that the prediction at the threshold of 5 minutes has the best fitness, and the prediction for the threshold of 13 minutes is also slightly good, while it is not so good in the case of 8 minutes. In fact, if checking the prediction for individuals, the accuracy of this model is still not enough to explain the individual behavior. But the trend line in Figure 6 infers that this result can reflect the behavior of people with specific individual characteristics at a given threshold of walking duration.

|  |
| --- |
|  |
| (a) |
|  |
| (b) |
|  |
| (c) |
| **Figure 6.** Trend Line of Prediction and Test Set: (a) is the trend line of prediction result and test set at the threshold of 5 minutes; (b) is the trend line of prediction result and test set at the threshold of 8 minutes, (c) is the trend line of prediction result and test set at the threshold of 13 minutes |

6. Discussion and Conclusion

This study regarded surveyed walking duration as the reflection of willingness on walking duration rather than the distance between departures and transit stations. From this perspective the present study investigated the prevalence of walking duration with respect to the individual characteristics. This study established the quantitative correlation between walking duration and individual characteristics, also provided a general perspective for estimating the factor influencing walking duration. The approach and results are expected can be used in estimating or planning the catchment area of public rail transit, which is also considered as the foundation for predicting transit ridership [24].

Viewing from the process of problem conversion, the data in this study is obtained from a factual investigation but not a willingness survey, which means that once people have made a decision of walking to rail transit stations, the walking duration is just representation of the location of departures and the walking speed. It is hard to say whether the individual characteristics affected the walking durations, maybe due to this reason few existing studies can explain the relevance between walking duration and individual characteristics quantitatively correctly [8,12,25]. In this study, the surveyed walking duration is regarded as reflection of the willingness on walking duration, thus it should relate to people's individual characteristics. On the base of assumption , the behavior that a passenger chose to walk to a station means this passenger can accept the walking duration from the departure to that station. From assumption and , if the passenger who accepted the given threshold of walking duration shows significant differences in individual characteristics, it means they have a different acceptability of walking duration with the others.

As explained above, this study examined the differences in individual characteristics of which people who accepted the given thresholds of walking duration. As the results, people with different individual characteristics shows different acceptability at each threshold of walking duration. According to the evaluation from the method of simple moving average, the model of 5 minutes’ threshold has a better explanatory ability, the model of 13 minutes' threshold is a little weaker, and the model of 8 minutes' threshold is not good. Possible expectations for these results are as follows. The selected thresholds of walking duration 5, 8, 13 minutes represent the lower boundary, mean value, and upper boundary of the main distribution of walking durations respectively. The commonly acceptable walking durations range from 5 to 13 minutes; therefore, it can be inferred that people who prefer walking less than 5 minutes and who can accept walking longer than 13 minutes may have significant features. However, since the mean value of walking duration is about 8 minutes, which means that most of the walking durations distributes around 8 minutes, for passengers there may be some randomness in making the decision of whether walking to rail transit stations. For the other threshold values near the mean value of 8 minutes, it also can be inferred that people may have some ambiguity in choosing whether to walk to stations or not. This explanation can also be confirmed by the result of valid features selection. There are 5 identical features in both thresholds of 5 and 8 minutes, which means people with those features are not sensitive to the threshold from 5 to 8 minutes.

Although the clear relationship between walking duration/distance and explanatory variables is rarely confirmed in the previous research, travel purpose was assumed to be one of the most important factors influencing walking duration [11,12]. Significance discrepancy with respect to travel purpose is also confirmed in some descriptive analysis [26,27]. According to the findings in this study, walking for private purpose, such as reaction and shopping, tends to be longer than that for work purpose, this finding is also supported by other research [11]. The dummy variable of peak hour is also assumed to be one of the important factor influencing walking duration to transit stations, however, significant effect was not confirmed in previous research [8,12]. This study found that passengers have an inclination of taking a shorter time to walk to stations during peak hours. A possible explanation for this result may be the tension in time, for which even if the acceptable walking duration has not changed for a passenger, the acceptable walking duration is supposed to become shorter. This explanation also can be mutually verified by the effect of travel purpose obtained in this study. Since peak hour commonly covers the trip with the purpose of work, passengers with the travel purpose of work are willing to get to stations faster as well. Moreover, this study also indicates that different age groups have different willingness on walking duration to transit stations, passengers aged from 25 to 44, which is in the range of productive age, tend to walk a shorter time to stations. This finding also can be supported by the effect of travel purpose and peak hour obtained in this study.

This study is the first step to explore the relationship between walking duration and individual characteristics, the findings in this study require further examination in terms of generality, and the way of applying the achievement need to be further discussed. The results of the random forest model perform not well on individual predictions. But for a group of people, it shows a good predictive ability, especially at the threshold of 5 minutes. In the next stage of this research, we plan to apply the results on predicting the willingness at a specific threshold of walking duration for a group of people. By using this prediction, for example, if knowing the individual characteristics of residents in a specific area locating from the rail transit station minutes, the general acceptability of walking to the station for the residents in this area should be predictive. Therefore, this prediction of willingness is expected to be used in planning the catchment area of rail transit stations or estimating the catchment area of existing stations.

**Author Contributions:** Qi Chen conceived and designed the research framework, analyzed the data, and developed the method. Go Higuchi provided support in data curation. Siting Chen wrote parts of the original draft and revised the whole paper. Shichen Zhao provided all the resource necessary during the research and had been controlling the research direction all along.

**Acknowledgments:** This study is finished during the author's doctoral period, the financial support from China Scholarship Council (CSC) is gratefully acknowledged.

**Conflicts of Interest:** The author declares no conflict of interest.

References

1. Kuby, M.; Barranda, A.; Upchurch, C. Factors influencing light-rail station boardings in the United States. Transp. Res. Part A Policy Pract. 2004, 38, 223–247, doi:10.1016/j.tra.2003.10.006.

2. Gutiérrez, J.; Cardozo, O. D.; García-Palomares, J. C. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. J. Transp. Geogr. 2011, 19, 1081–1092, doi:10.1016/j.jtrangeo.2011.05.004.

3. Cardozo, O. D.; García-Palomares, J. C.; Gutiérrez, J. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Appl. Geogr. 2012, 34, 548–558, doi:10.1016/j.apgeog.2012.01.005.

4. Zhao, J.; Deng, W.; Song, Y.; Zhu, Y. What influences Metro station ridership in China? Insights from Nanjing. Cities 2013, 35, 114–124, doi:10.1016/j.cities.2013.07.002.

5. Frank, L.; Pivo, G. Impacts of mixed use and density on utilization of three modes of travel: Single occupant vehicle, transit, and walking. Transp. Res. Rec. 1994, 1466, 44–52, doi:10.1073/pnas.1100480108.

6. Waterson, B. J.; Rajbhandari, B.; Hounsell, N. B. Simulating the impacts of strong bus priority measures. J. Transp. Eng. 2003, 129, 642–647, doi:10.1061/(ASCE)0733-947X(2003)129.

7. Besser, L. M.; Dannenberg, A. L. Walking to public transit: steps to help meet physical activity recommendations. Am. J. Prev. Med. 2005, 29, 273–280, doi:10.1016/j.ampre.2005.06.010.

8. Daniels, R.; Mulley, C. Explaining walking distance to public transport: The dominance of public transport supply. J. Transp. Land Use 2013, 6, 5–20, doi:10.5198/jtlu.v6i2.308.

9. Guerra, E.; Cervero, R.; Tischler, D. Half-mile circle: Does it best represent transit station catchments? Transp. Res. Rec. J. Transp. Res. Board 2012, 2276, 101–109, doi:10.3141/2276-12.

10. Larsen, J.; El-Geneidy, A.; Yasmin, F. Beyond the quarter mile: re-examining travel distances by active transportation. Can. J. Urban Res. 2010, 19, 70–88.

11. Yang, Y.; Diez-Roux, A. V Walking distance by trip purpose and population subgroups. Am. J. Prev. Med. 2012, 43, 11–19, doi:10.1016/j.amepre.2012.03.015.

12. Krygsman, S.; Dijst, M.; Arentze, T. Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. Transp. Policy 2004, 11, 265–275, doi:10.1016/j.tranpol.2003.12.001.

13. McCormack, G. R.; Cerin, E.; Leslie, E.; Du Toit, L.; Owen, N. Objective versus perceived walking distances to destinations: correspondence and predictive validity. Environ. Behav. 2008, 40, 401–425.

14. Marchetti, C. Anthropological invariants in travel behavior. Technol. Forecast. Soc. Change 1994, 47, 75–88, doi:10.1016/0040-1625(94)90041-8.

15. Weinstein Agrawal, A.; Schlossberg, M.; Irvin, K. How far, by which route and why? A spatial analysis of pedestrian preference. J. urban Des. 2008, 13, 81–98, doi:10.1080/13574800701804074.

16. Ho, T. K. Random decision forests. In Document analysis and recognition, 1995., proceedings of the third international conference on; 1995; Vol. 1, pp. 278–282.

17. Ho, T. K. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 832–844, doi:10.1109/34.709601.

18. Murray, A. T.; Davis, R.; Stimson, R. J.; Ferreira, L. Public transportation access. Transp. Res. Part D Transp. Environ. 1998, 3, 319–328.

19. O’Sullivan, S.; Morrall, J. Walking distances to and from light-rail transit stations. Transp. Res. Rec. J. Transp. Res. board 1996, 1538, 19–26, doi:10.3141/1538-03.

20. Keijer, M.; Rietveld, P. How do people get to the railway station? The Dutch experience. Transp. Plan. Technol. 2000, 23, 215–235, doi:10.1080/03081060008717650.

21. Zhao, F.; Chow, L.-F.; Li, M.-T.; Ubaka, I.; Gan, A. Forecasting transit walk accessibility: Regression model alternative to buffer method. Transp. Res. Rec. J. Transp. Res. Board 2003, 1835, 34–41, doi:10.3141/1835-05.

22. Alshalalfah, B.; Shalaby, A. S. Case study: Relationship of walk access distance to transit with service, travel, and personal characteristics. J. urban Plan. Dev. 2007, 133, 114–118, doi:10.1061/(ASCE)0733-9488(2007)133:2(114).

23. Ohannon, R. W. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. Age Ageing 1997, 26, 15–19, doi:10.1093/ageing/26.1.15.

24. Guerra, E.; Cervero, R. Is a Half-Mile Circle the Right Standard for TODs? ACCESS Mag. 2013.

25. Levinson, D. M.; Kumar, A. Density and the journey to work. Growth Change 1997, 28, 147–172, doi:10.1111/j.1468-2257.1997.tb00768.x.

26. Burke, M.; Brown, A. L. Distances people walk for transport. Road Transp. Res. 2007, 16, 16–29.

27. Hoback, A.; Anderson, S.; Dutta, U. True walking distance to transit. Transp. Plan. Technol. 2008, 31, 681–692, doi:10.1080/03081060802492785.

H:\documents\layout\new template June 2014\figures\CC-BY logo original v1.wmf© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).