Passengers' Hyperpath Choice Behavior Observation on Transit Network Using Smartcard Data

Seham M.HASSAN[1], Chen, SHENG-YU[2], Jan-Dirk SCHMÖCKER[3], Fumitaka KURAUCHI[4] and Daisuke FUKUDA[5]

[1] Master course student, Dept of Civil Eng., Gifu University
(1-1 Yanagido, Gifu City 501-1193,Japan)
E-mail:sehassan2007@yahoo.com

[2]Alumni of National Central University, Taiwan
(No.300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan)
Email:93302064@cc.ncu.edu.tw

[3]Associate Professor, Ph.D, Dept of Urban Management, Kyoto University
(Kyotodaigakukatsura, Kyoto, 615-8540, Japan)
E-mail:schmoecker@trans.kuciv.kyoto-u.ac.jp

[4]Associate Professor, Dr. of Eng., Dept of Civil Eng., Gifu University
(1-1 Yanagido, Gifu City 501-1193,Japan)
E-mail:kurauchi@gifu-u.ac.jp

[5]Associate Professor, Dr. of Eng., Dept of Civil Engineering, Tokyo Institute of Technology
(M1-11, 2-12-1, O-okayama, Meguro-ku,Tokyo 152-8552, Japan)
E-mail:fukuda@plan.cv.titech.ac.jp

A hyperpath can be defined as a set of attractive lines identified by the passenger for each stop, each of which might be the optimal one from the stop, depending on lines arrival time, frequency, cost etc. The concept of the hyperpath has been used for most of transit assignment models as a fundamental assumption, although the passengers' behavior on the real transit network has not been well explored. This research uses time series smart card data from London to investigate whether passengers follow such proposed hyperpaths or not. The analysis is based on *n*-step Markov models and proposes that the variations in routes taken by passengers who supposedly travel between the same OD pair every morning over several days should reflect the set of paths included in an (optimal) hyperpath. Our hypothesis is that a large variation in bus lines over days indicates a complex hyperpath whereas a passenger who takes the same line every morning does not consider many alternatives. Our results suggest that there is some variation in routes chosen, possibly in accordance with the theory of hyperpaths in networks with uncertainty.

Key Words: *Hyperpaths, Oyster Card data, Travel behavior, Markov model*

1. INTRODUCTION

It is generally assumed that on transit networks travelers try to minimize their expected travel time consisting of waiting time, on-board time as well as potentially other factors such as fare, crowding or seat availability by selecting a hyperpath. A hyperpath can be defined as a set of attractive lines identified by the passenger for each stop, each of which might be the optimal one from the stop, depending on lines arrival time, frequency, cost etc. In networks with few uncertainties, e.g. regular arrival times, low congestion, this set of services will be smaller as passengers can better estimate whether it is advantageous to let slow services pass in order to wait for the faster service that might arrive soon. This behavioral assumption has led to a fairly large set of literature.

Several transit network design models that aim at optimizing the frequencies of the lines have been proposed[1)]. Transit assignment has been studied either as a separate problem (e.g. see Andreasson[2)]), or as a sub problem of more complex models, such as transit network design (e.g. Mandl[3)]), or multimodal network equilibrium (Florian[4)], and Florian et al.[5)]). Most of the algorithms that have been proposed in the past may be classified as heuristic approaches to the problem. These algorithms are variants of assignment procedures used for private car traffic on road networks (such as shortest path, stochastic multipath assignment) that are modified to reflect the waiting time phenomenon inherent to transit networks. Spiess et al.[6] replace the simplistic route choice models by a transit assignment model leading to more realistic transit network design models. They described a model for the transit assignment problem with a fixed set of transit lines. The traveler chooses the strategy that allows him or her to reach his or her destination at minimum expected cost. In recent approaches[7], the previous assumptions are refined; partly to reflect the increasing amount of information passengers obtain during their journeys. Another formulation is proposed by Nguyen et al.[8]. They introduced the concept of hyperpaths and formulate an equilibrium model similar to that used for congested road networks [9]. In both models, Spiess[9] and Nguyen et al.[8], waiting times are considered constant and independent of volumes.

The purpose of this research is to understand whether passengers indeed follow such proposed hyperpaths or whether habits and other factors would dominate routing decisions leading to less (or more) complex hyperpaths than those proposed in the literature. Observing hyperpaths is, however, difficult. One would have to understand which (unchosen) routing options the traveler considers. As a first step this analysis assumes that the variations in routes taken by regular commuters during their morning journey over several days should reflect the set of paths considered by hyperpath travelers in networks with uncertain vehicle arrival times. Our hypothesis is that a large variation in routes over days indicates a complex hyperpaths whereas a traveler who takes the same route every morning does not consider many alternatives leading to a simple hyperpaths.

The reminder of this paper is structured as follows. The next section describes the data used for this research and the preprocessing. Section 3 describes the Markov analysis used to analyze stability in route choice over days and shows some initial results. Section 4 describes how overlapping between routes is considered and how this changes the results. Section 5 discusses the findings of this paper.

2. DATA DESCRIPTION AND PREPARATION

Through Transport for London (T*f*L) smart card data from London's public transport network, commonly referred to as "Oyster card", has been obtained. It is believed that London is a good case study for our analysis because of three reasons. First, the public transport network is large and dense offering passengers a large number of route choices. Second, public transport services are operated frequency based. This is, even though there might be an internal timetable within T*f*L, in many cases passengers only find information about service frequencies at bus stops for frequent services during peak hour. Third, service reliability in London is not as high as in many other cities with smart card systems. All three reasons should encourage passengers to consider in many cases fairly complex hyperpaths.

The Oyster smart card system is implemented in London's bus, tube, tram, DLR as well as parts of its commuter rail system[10]. Smart card data are convenient for travelers, operators as well as analysts, allowing for example to conduct the analysis on which this paper is based. Kusakabe, et al.[11] describe in more detail possibilities as well as limitations of smart card data. There is an average of 6.3 Million Oyster card swipes recorded in our data set per day. Cardholders traveling by bus do only have to swipe when boarding a bus. Travelers on all track bound modes though have to swipe when boarding as well as when alighting. For the purpose of our analysis on path choice decision this additional alighting record is obviously advantageous. However, interchanges between tube lines are not captured by Oyster card, whereas the bus data do record the route number taken. Furthermore, the bus network offers users far more routing options and potentially complex hyperpaths with several bus routes departing from the same stop. Therefore our initial analysis focuses on bus records only.

We obtained two weeks of Oyster card's data for the period 08 Nov - 22 Nov 2007. Due to the size of our data set as well as some incomplete records significant preprocessing is required. As Kusakabe et al.[11] also noted this preprocessing effort required for the use of smart card data can be very substantial. Firstly we reduced the dataset to only that information relevant to our analysis. These are

- Card number, to identify the same traveler over several days,
- Route ID, to identify the bus route,
- Boarding time.

We further kept the information (bus stop) boarding location in our database, however, unfortunately the boarding location recorded on the Oyster card are not reliable. (This is because in 2007 the bus is not yet connected to the bus GPS system.) Note further that Oyster card does not record the exact bus the passenger is boarding. Clearly these are limitations to our study, but we believe to partly overcome these by reducing our sample to only those travelers who use a bus every day of the 10 week days in our sample before 9.30 am, meaning that we are likely to pick up only regular morning commuters. Any Oyster card holder who is not using a bus before 9.30am at all ten weekdays is excluded from our sample. Furthermore, we only select the first boarding of the day, meaning that it is likely that the person is starting the journey each day from the same location, his home. With these stringent conditions we reduce our sample size to 17302 regular bus commuters. Furthermore, we reduce our sample to those commuters whose hyperpath presumably only consists of bus route options from their home location.

3. INITIAL *N*-STEP MARKOV ANALYSIS

To analyze the consistency in route choice over days, we adopted an *n*-Step Markov model. The choice of route on day *d* is assumed to depend on the choices on *n* previous days. As we are not interested in which specific route the passenger is taking, but only in whether the traveler is taking the same or a different route the choice on the first day is generally abbreviated with bus A in the following. On the next day the passenger has then the choice to take the same bus A or a different bus B. If the passengers took two different buses on the first two days, on the third day he/she then has a choice between buses A, B, C and so on.

In the first analysis step we assume n=2. Our choice of independent days is taken as the previous day as well as the same weekday during the previous week. The letters follow chronological order of choices from left to right. Therefore, the two letters before the underline indicate the routes taken on previous days and the last letter indicates the route chosen on the predicted day. For example *AA_A* indicates that the traveler is taking the same route on all three days, *AB_B* indicates that the traveler is taking the same bus as yesterday, but that he took a different bus route on the same weekday on the previous week. *AB_C* indicates that different buses are taken each day. The 3-Step case can also be calculated in the same way, assuming e.g. the day before yesterday as an additional independent variable. The general form of the probability that a person *i* will choose line *j* on day *d* can be described as follows;

$$p_{ij}^{d} = \begin{cases} \Pr(A \mid AA) & \text{if } j_d = j_{d-1}, j_{d-1} = j_{d-7} \\ \dfrac{\Pr(C \mid AA)}{J_i - 1} & \text{if } j_d \neq j_{d-1}, j_{d-1} = j_{d-7} \\ \Pr(A \mid AB) & \text{if } j_d = j_{d-1}, j_{d-1} \neq j_{d-7} \\ \Pr(B \mid AB) & \text{if } j_d = j_{d-7}, j_{d-1} \neq j_{d-7} \\ \dfrac{\Pr(C \mid AB)}{J_i - 2} & \text{if } j_d \neq j_{d-1}, j_d \neq j_{d-7}, j_{d-1} \neq j_{d-7} \end{cases}$$

Where:

i	:	person,
j	:	transit line choice,
j_d	:	transit line choice on the *d*th day,
p_{ij}^{d}	:	probability that a person *i* chooses transit *j* on the *d*th day,
J_i	:	number of available transit lines for person *i*.

Our hypothesis is that if we expect some route variations then hyperpaths may exist (i.e. a large part of the route choice variation may be due to common lines that are part of the travelers hyperpaths).

Fig. 1 describes the probability of each choice for 2-step Markov model, suggesting that there is considerable variation in routes chosen, and indicates that a large amount of commuters change route at least on some days. It also shows that only around 23% choose the same bus every day whereas around 20% choose a different bus route on all three days. The reminding percentage of passengers chooses a different bus on at least one out of the three days. **Fig.2** shows the results for n=3 where the independent days are taken as the day before, two days before and the same weekday in the week before. The percentage of commuters taking every day the same bus now reduces to less than 17%. Note, however, that the percentage of commuters taken a different bus every day reduces even further to below 6%. The reminding percentage of passenger chooses the same bus on at least two out of the four days.

Both figures further indicate that the day of the week for which the route choice is predicted does not appear to have a significant influence on the results. Furthermore, it cannot be observed that either the previous day or the same day of the last week has a more significant influence on the prediction. These results seem reasonable, given the selection of our sample (first bus chosen from home of regular commuters). Our results might suggest that there is indeed some random variation in routes chosen, possibly in accordance with the theory of hyperpaths in networks with uncertainty.

4. CONSIDERATION OF OVERLAPPING ROUTES

In order to understand whether the variation in chosen routes observed in **Figs 1** and **2** is indeed due to passengers traveling on hyperpaths or whether this is due to other reasons, overlap of routes is considered in this section. Our hypothesis is that we predict would the variation in route choice to disproportionally decrease in case line overlap is

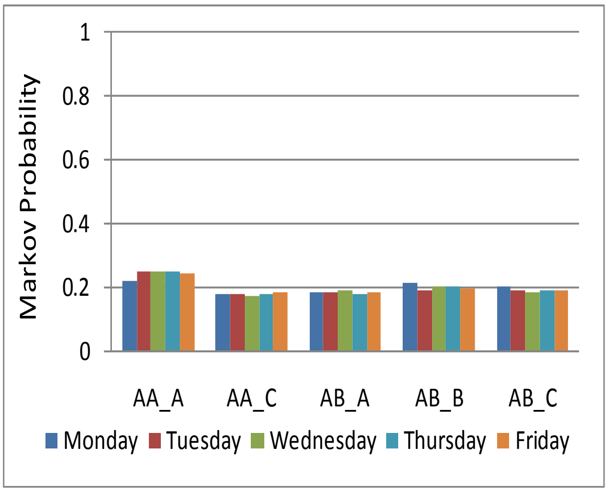


Fig.1 2-Step results for route predication during the 2nd week

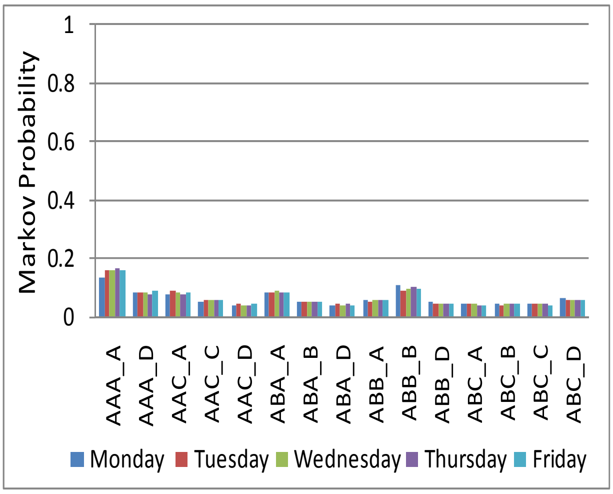


Fig.2 3-Step results for route predication during the 2nd week

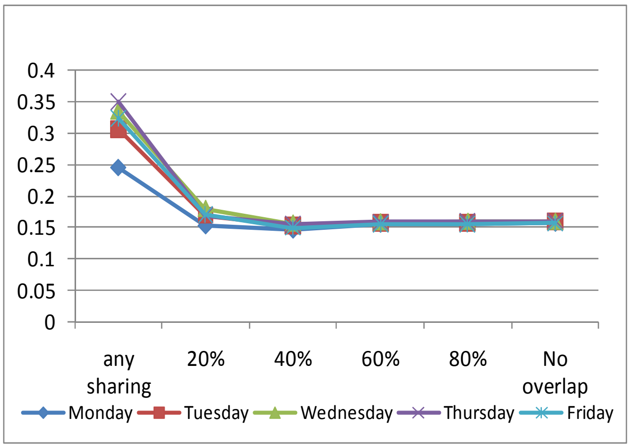


Fig.3 Good of fit index (ρ) for different overlapping thresholds S

Table 1 Good of fit index for different thresholds S

	0% (any)	20%	40%	60%	80%	100% (none)
Mon	0.245	0.153	0.146	0.155	0.157	0.157
Tues	0.306	0.170	0.154	0.158	0.158	0.159
Wed	0.336	0.180	0.156	0.158	0.159	0.159
Thrs	0.350	0.170	0.156	0.159	0.160	0.161
Fri	0.324	0.171	0.150	0.155	0.156	0.157

considered. This indicates that a large part of the route choice variation is indeed due to common lines that are part of the travelers' hyperpaths.

As explained before though, unfortunately our data set does not allow us to induce the home bus stop of the respondent. Therefore we cannot identify directly whether a traveler took a different route B on a second day because it is part of the same hyperpath or because of different reasons. As an approximation we can only identify the degree to which the routes overlap. We therefore define p_{xy} as the percentage of shared stops on the routes of two bus lines x and y.

If the percentage exceeds a predefined threshold S these two lines are considered as the same line as it is presumed likely that passengers could take both lines from their home location. This means that the smaller S the smaller the set of lines. S=0 would mean that the traveler always faces only the option of 1 line, whereas S=1 leads to the results identically to those shown in the previous section.

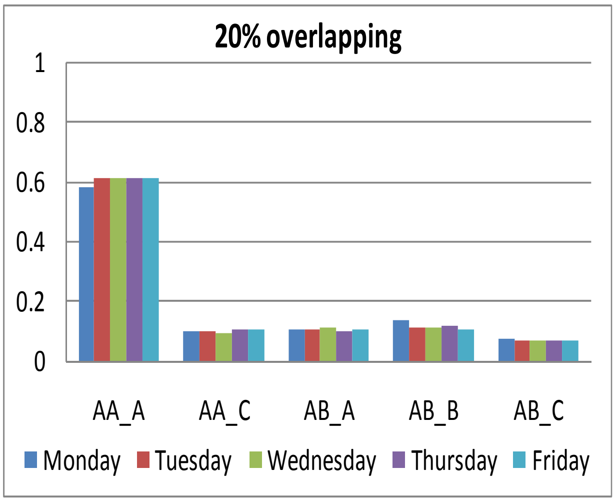
Tests were carried out with different thresholds S for 2-step Markov models for prediction of route choice during the second week. **Fig.3** shows the goodness of fit index (ρ) for different overlapping thresholds S during week days, which are calculated with the following equation.

$$\rho=1-LL\,(0)/LL\,(\text{model}) \qquad (1)$$

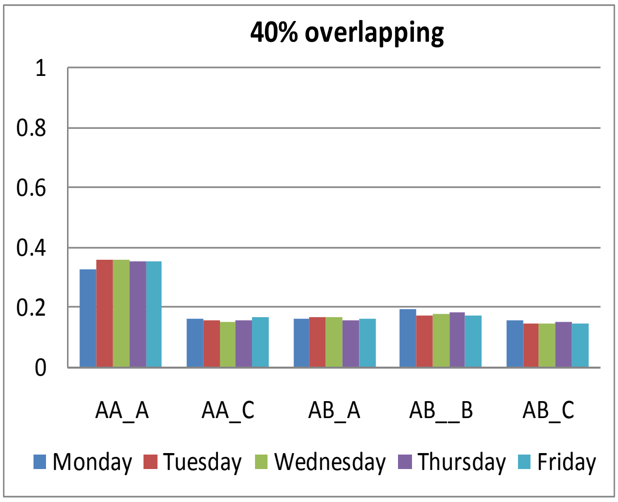
It is concluded from **Fig.3** that considering overlap is important to increase the model fit of the predicted model. The likelihood ratio index (ρ) improved significantly for low overlapping thresholds S (i.e. 20% is better than 40% and so on) but no improvement, compared to ignoring overlapping, can be observed for S>40%. **Table 1** shows the goodness of fit indices for the 2-step Markov analysis and **Fig.4(a)-(d)** illustrates the results of the 2-Step Markov model for route predication during the 2nd week for different overlapping thresholds S.

5. DISCUSSION

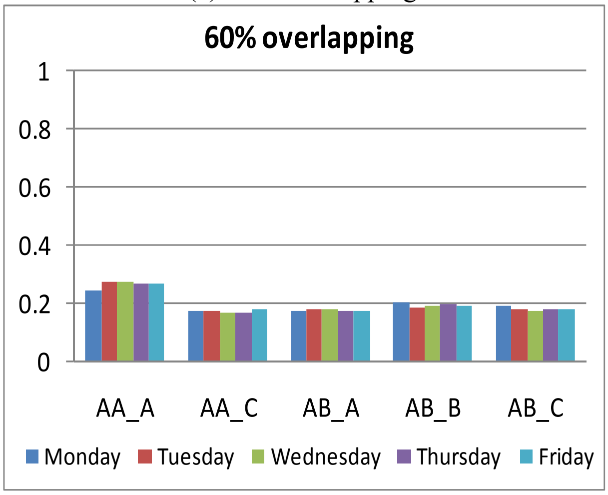
Comparing **Fig.1** and **Fig.4**, we have an evidence that the variation in route choice is decreased in case line overlap is considered. This may indicate that a large part of the route choice variation is indeed due



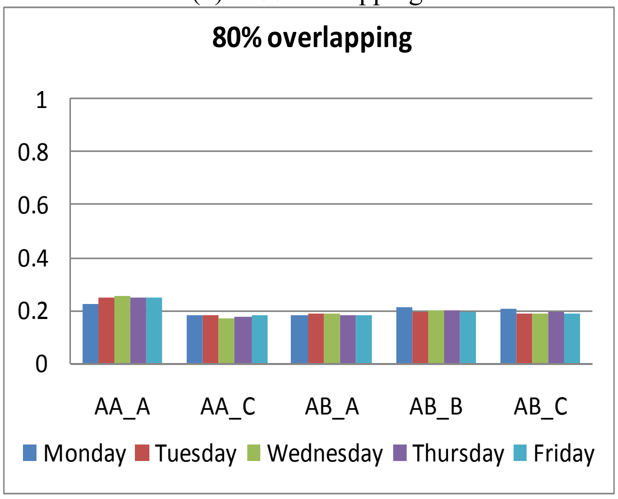
(a) 20% overlapping



(b) 40% overlapping



(c) 60% overlapping



(d) 80% overlapping

Fig.4 2-Step results with S=0.2, S=0.4, S=0.6, and S=0.8

to common lines that are part of the travelers' hyperpaths or at least some route variation is due to overlap and possibly hyperpaths. $S > 60\%$ reduces the significance of considering overlapping as there remains significant variation in route choice prediction. Only for S <0.4, the variation in route choice decreases compared to route choice prediction without considering overlapping.

We conclude that Markov models can be used to analyze the consistency in travelers' route choice behavior observed with time series Smart Card data. The result shows that the day of the week for which the route choice is predicted does not appear to have a significant influence on the results. Furthermore, initial results might suggest that there is some random variation in routes chosen, possibly in accordance with the theory of hyperpaths in networks with uncertainty. Overlap of routes is considered to understand whether this is indeed due to passengers travelling on hyperpaths or whether this is due to other reasons. It is found that the variation in route choice is decreased in case line overlap is considered. This may indicate that a large part of the route choice variation is indeed due to common lines that are part of the travelers' hyperpaths or at least some route variation is due to overlap and possibly hyperpath.

ACKNOWLEDGEMENT

This research has been carried out in collaboration with the DynCapCon project headed by Prof. Michael G. H. Bell, Imperial College London and Dr. Achille Fonzone (now Edinburgh Napier University). Transport for London kindly provided us with the Oystercard data. We thank to all the people and organizations for supporting this research.

REFERENCES

1) Fearnside K. and Draper D. P.: Public transport assignment-A new approach. *Traffic Engng. Conrrol*, pp. 298-299, 1971.

2) Andreasson I.: A Method for the Analysis of Transit Networks. *Proceedings of the 2nd European Congress on Operations Research* (Edited by Mark Roubens). North Holland, Amsterdam, 1976.
3) Mandl C. E. : Evaluation and Optimization of Urban Public Transportation Networks. *Presented at the 3rd European Congress on Operations Research*. Amsterdam, Netherlands, 1979.
4) Florian M. A :traffic equilibrium model of traffic by car and public transit modes, *Transportation Research* Vol. 10, pp. 166-179, 1977.
5) Florian M. and Spiess H.: On two mode choice/assignment models. *Trunspn. Sci.* Vol. 17, pp. 32-47, 1983.
6) Spiess, H and Florian, M.: Optimal Strategies: A new assignment model for transit networks. *Transportation Research*, Vol. 23B(2), PP. 83-102, 1989.
7) Nökel, K. and Wekeck S.: Boarding and alighting in frequency-based transit assignment. *Paper presented at 88th Annual Transportation Research Board Meeting*, Washington D.C., 2009.
8) Ngyuen S. and Pallotino S.:Equilibrium Traffic Assignment For Large Scale Transit Networks, Eur. *J.Opnl.Res.*,1988.
9) Spiess H.: on optimal route choice strategies in transit networks, Pub.286,*Center de recherché sur les Transports*,Universite de Montreal,1983.
10) Transport for London : http://www.tfl.gov.uk/tickets/14836.aspx
11) Kusakabe, T., Iryo, T., & Asakura, Y.: Estimation method for railway passengers' train choice behavior with smart card transaction data. *Journal of Transportation*, Vol. 37, (5), pp. 731-740, 2010.