Empirical Analysis on Passengers' Hyperpath Construction by Smart Card Data

Fumitaka KURAUCHI	Gifu University
Jan-Dirk SCHMÖCKER	Kyoto University
Hiroshi SHIMAMOTO	Kyoto University
Seham M. HASSAN	Gifu University

Paper presented at CASPT2012, Santiago, 25th, July, 2012

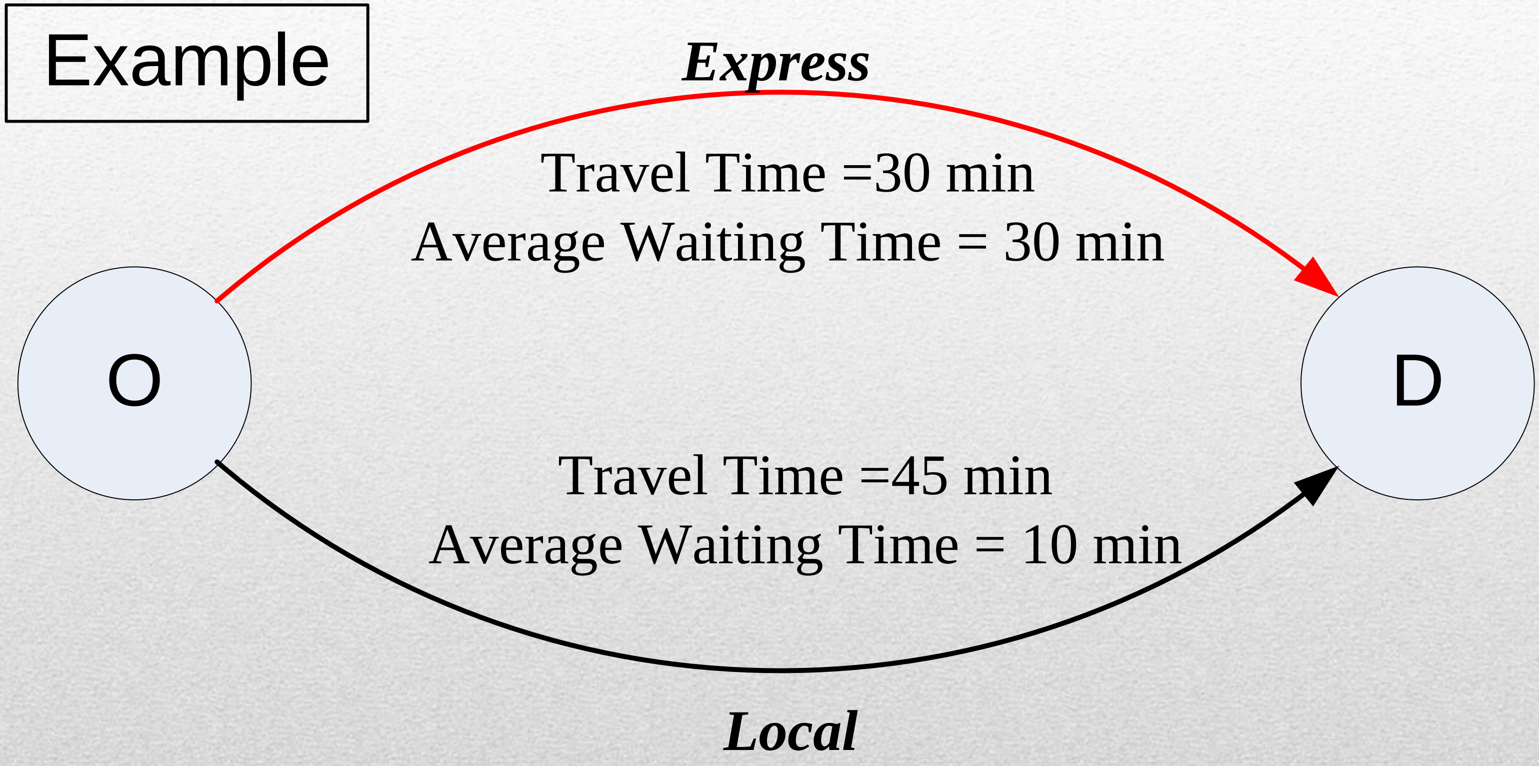
Introduction

- Public transit planning is getting more important in terms of energy saving and green gas emission mitigation.
  - Good model for reproducing passenger flows is required.
- In transit assignment modeling, it is often assumed that passengers choose strategies which possibly include complex path sets to minimise their expected travel time.
- Only relatively small emphasis on the estimation of passengers' choice sets to discuss whether line choices are "random" based on transit service attributes such as which line from this choice set arrives first.

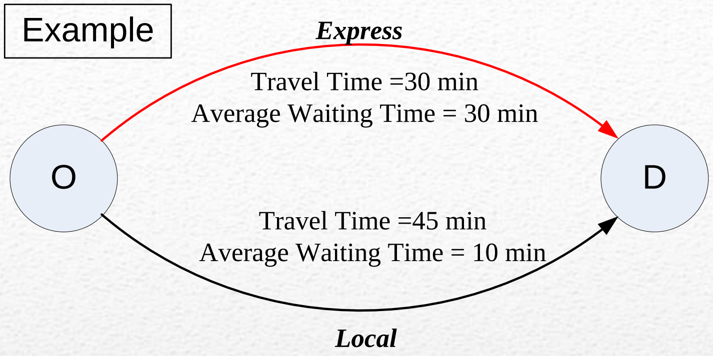
Definition of Common Lines Problem

- When there are several lines sharing the same platform and all of them can be used to reach the destination, then the expected travel time might decrease by boarding the train whichever arrives first.
- Path of the shortest expected travel time is not the elementary path but the set of elementary paths (hyperpath by Spiess and Florian, 1987)

Simple Example

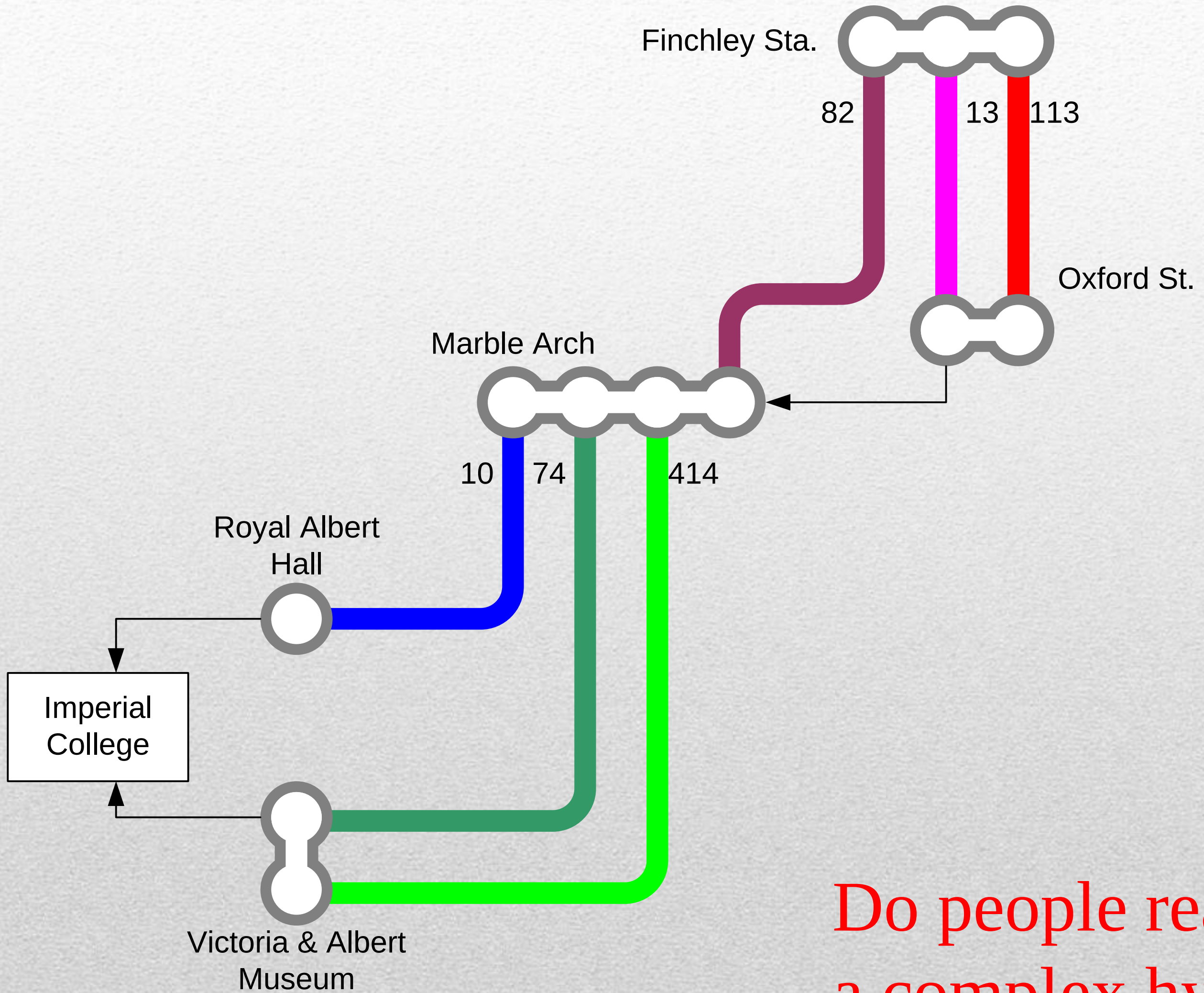


Path of Shortest Time



- Path of shortest travel time
  - Express : 30 min
  - Local : 45 min
- Path of shortest time considering average waiting time
  - Express : 60 min (30 + 30)
  - Local : 55 min (45 + 10)
- Why not getting on Express when it comes first?
  - If a passenger gets on the train whichever comes first..
    - Travel time :0.25 x 30 + 0.75 x 45 = 41.25 min
    - Waiting time :1/(1/10+1/30)=30/4=7.5 min
    - Expected travelling time :48.75 min

The Optimal Hyperpath Can Be Complex…



Do people really construct such a complex hyperpath!?

Objectives

- Explore the passenger behaviour by using London Oystercard data, especially whether passengers construct hyperpaths or not.

We want to fill the gap between 'optimal' path choice assumed in assignment model and path choice observations.

- Because of the data limitation, we discuss here about the 'regularity' of the first trip every day, and try to find indirect evidence for hyperpaths being used.

London PT network characteristics

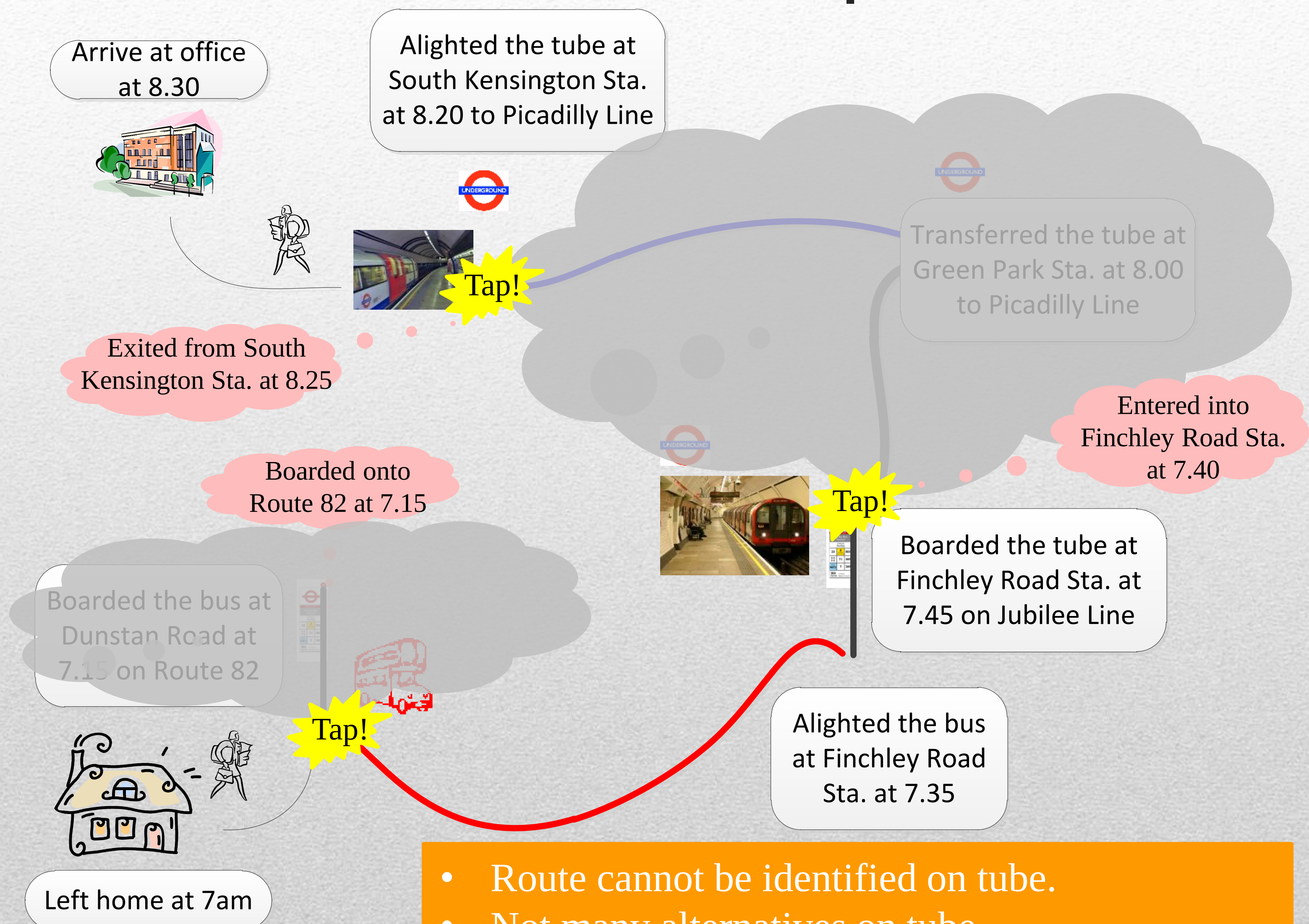
- Oystercard is a plastic smartcard one can use on bus, tube and (partially) train in London.
- London's public Transport network is large and dense offering passengers a large number of route choices.
- Public Transport services are operated frequency based.
- Service reliability in London is not as high as in many other cities with smart card systems.

http://www.tfl.gov.uk/tickets/14836.aspx

Data Description

- Oyster card dataset
  - Tube : In + Out Station and time+ card ID
  - Bus : Route + Boarding Time + card ID
- Date
  - 08 Nov - 22 Nov 2007
- Data Limitation
  - Underground is not dense enough to construct hyperpath, and the route cannot be observed.
  - The accuracy of records of bus boarding location is low, because the bus is not yet connected to the bus GPS system (iBus).
  - Oystercard does not record the exact bus ID, just bus line number.
  - Alighting point not known because there is no tap out (bus).
- Data extraction:
  - Pick up the traveller who use bus on all 10 working days before 9.30 am (i.e. regular morning commuters).
  - Select the first boarding line of the day(assume: the same location, home).
  - 17,302 regular bus commutes in our database.

Data Limitation - Example



- Route cannot be identified on tube.
- Not many alternatives on tube.
- Boarding/alighting bus stops are unknown.

Methodology

- n-step Markov analysis
  - The choice of route on day *d* is assumed to depend on the choices on *n* previous days.
  - Likelihood ratio is used to check the goodness of model fit.
  - Expectation
    - Used line every day may vary depending on the bus arrival, if travellers construct hyperpath,.
    - In such case, not many travellers stick on the single line.
- Overlapping analysis
  - Routes are 'bundled' accordingly to the degree of overlapping for each route, and Markov analysis is carried out by the bundles.
  - Improvement of the goodness of model fit may be due to the hyperpath construction.

Markov Analysis

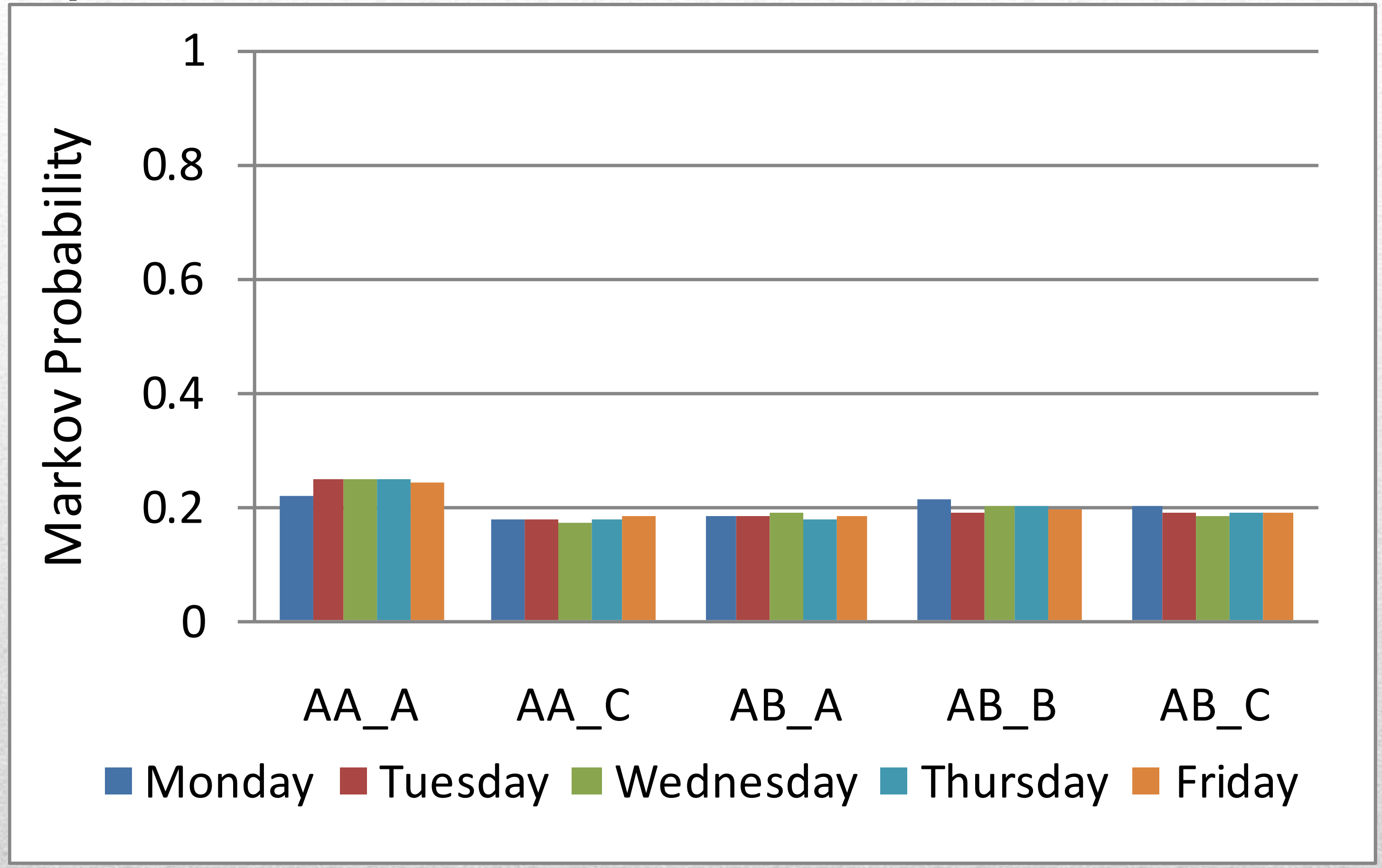
- The choice of route on day *d* is assumed to depend on the choices on n(n=2 or 3) previous days (previous day and same weekday in previous week)

Example of 2-step Markov Model

Same day in last week	Previous day	Predicted day = same day last week	Predicted day =yesterday	Predicted day ≠same day last week ≠ yesterday
$j_{d\text{-}7}$	$j_{d\text{-}1}$	$\Pr(j_d = j_{d\text{-}7})$	$\Pr(j_d = j_{d\text{-}1})$	$\Pr(j_d \neq j_{d\text{-}7}, j_d \neq j_{d\text{-}1})$
A	*A*	Pr(*AA_A*)	-	Pr(*AA_C*)
A	*B*	Pr(*AB_A*)	Pr(*AB_B*)	Pr(*AB_C*)

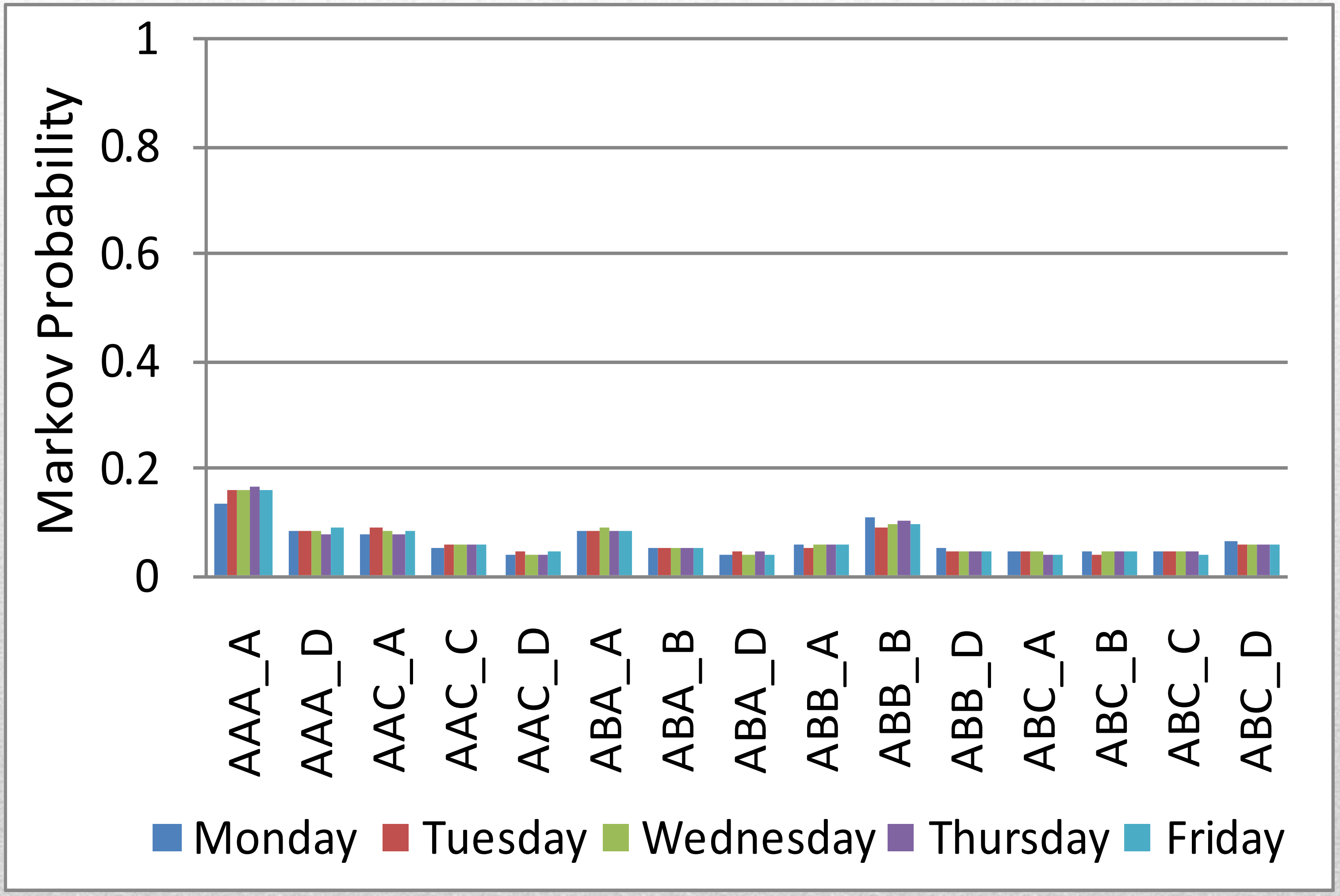
- These probability are calculated based on oystercard data.
- The stability of the transit route choice can be analysed by them.

2-Step Results for Route Choice Prediction



- Large variation in routes chosen.
- The day of the week for which the route choice is predicted does not have an influence on the results.

3-Step Results for Route Choice Prediction

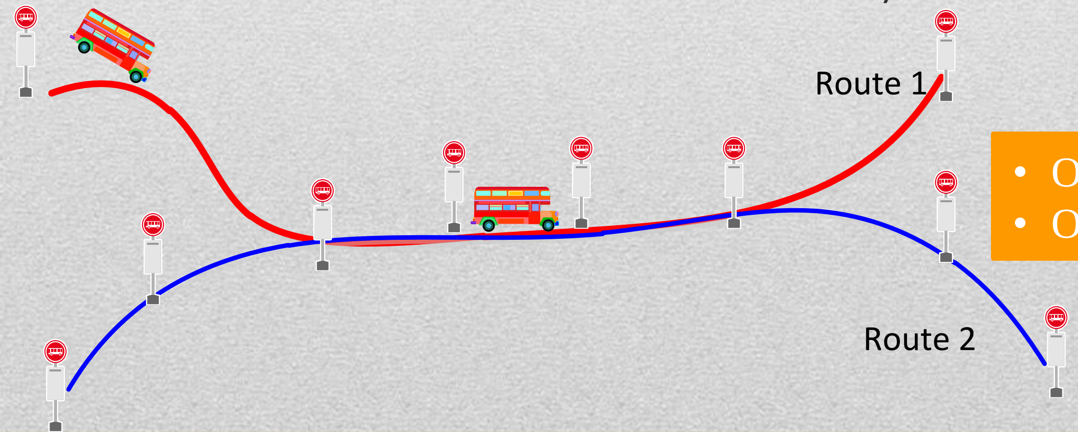


Discussion on the Initial Results

- A large variation in routes over days may indicate a complex hyperpaths whereas a traveller who takes the same route every morning does not consider many alternatives,
- There seems to be some random variation in routes chosen, possibly in accordance with the theory of hyperpaths in networks with uncertainty.
- To understand whether the observed variation in chosen routes is due to passengers travelling on hyperpaths or other reason, overlap in routes is considered to overcome the data limitation.
- If the variation in route choice decrease in the case if line overlap is considered, a large part of the route choice variation can be explained by the travellers' hyperpath construction.

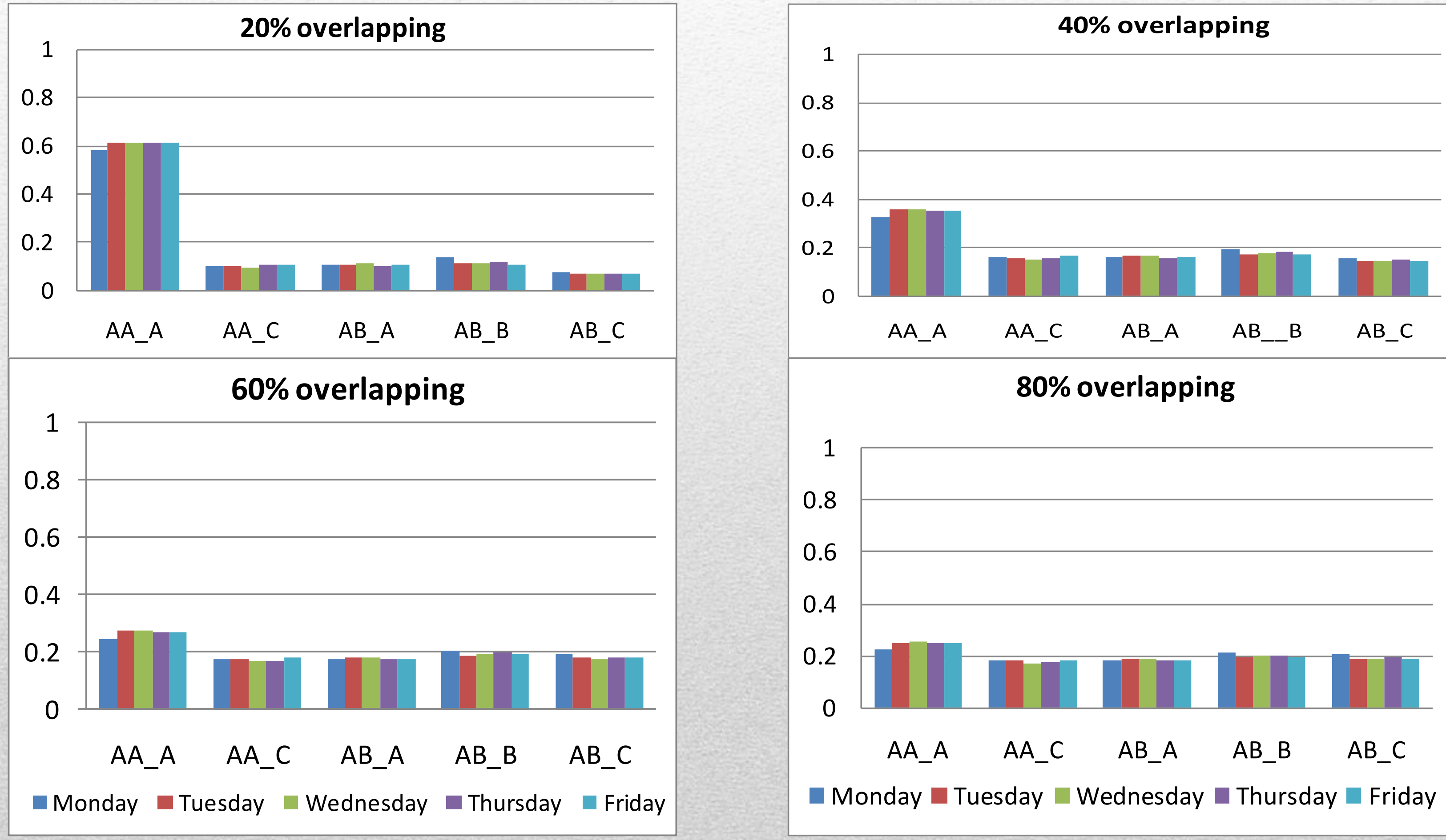
Consideration of Overlapping

- Since we don't have an exact boarding/alighting bus stops…
  - We cannot directly know whether the passenger can move to the same bus stop by other lines.
  - As an approximation, we define the overlapping rate as the percentage of the number of shared bus stops.
  - If overlapping rate is large, the lines are likely to run on the similar route, and thus they are likely to construct the hyperpath.
- If the overlapping rate exceeds a predefined threshold S these two lines are considered as the same line (the bundle of the lines).
  - S=0% means the lines are treated as a bundle if they share at least one bus stop.
  - S=100% means that all lines are regarded as independent (initial result).
  - Smaller S means the lines are likely to be 'bundled'.



- Overlapping rate of route 1 with route 2 is 0.66.
- Overlapping rate of route 2 with route 1 is 0.50.

Route Choice Prediction Considering Overlaps



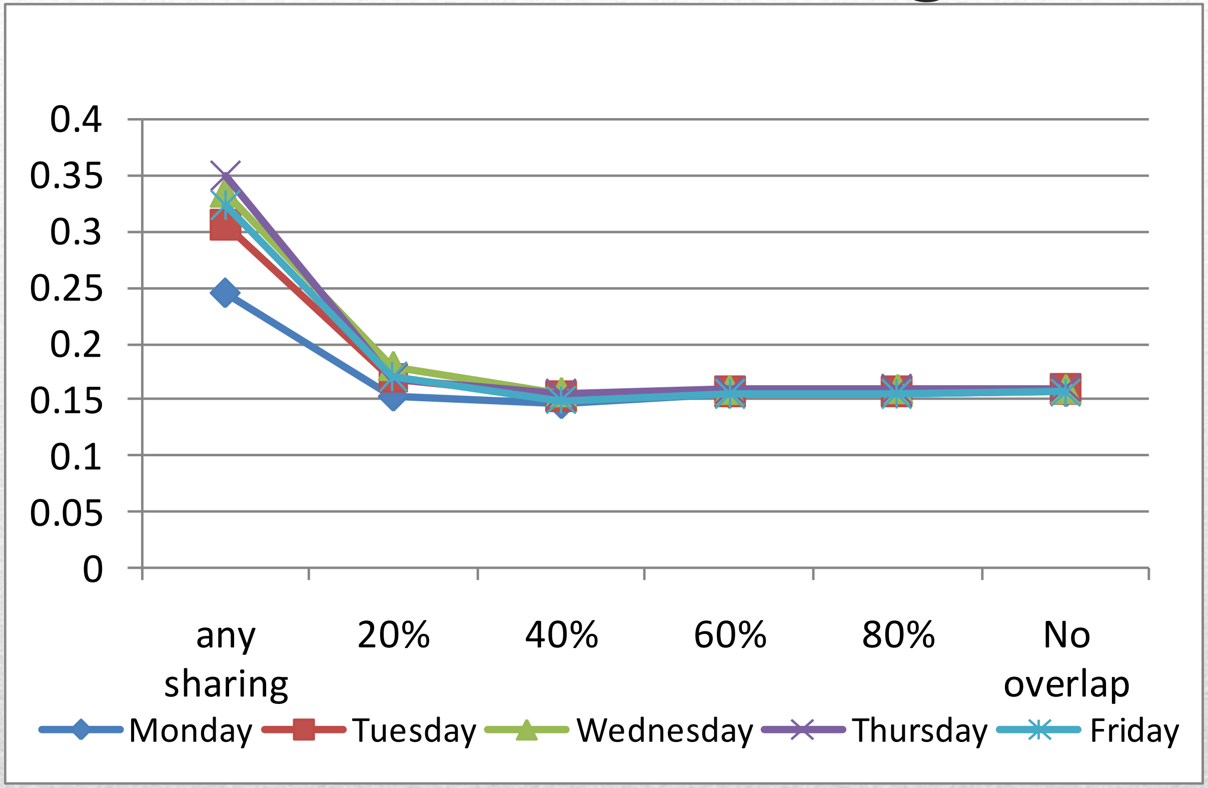
- Obviously the variation in route choice is decreased in case line overlap is considered.

Likelihood Analysis

- Likelihood ratio is calculated to check the model fitness.
- The goodness of model fit index (ρ) is determined for different overlapping thresholds S during week days according the following equation.
- The number of cases (alternatives) considered to avoid overestimation.

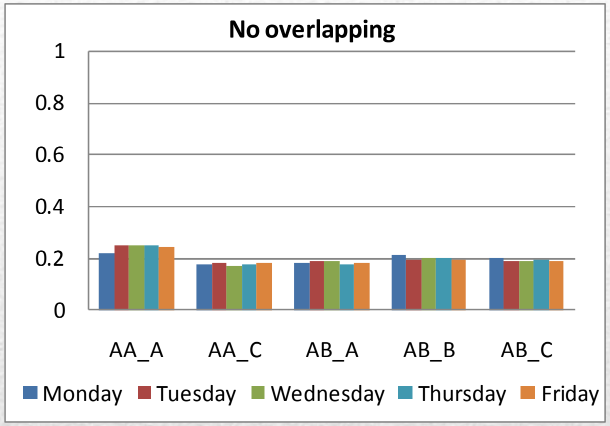
$$\rho = 1 - LL(0)/LL(\text{model})$$

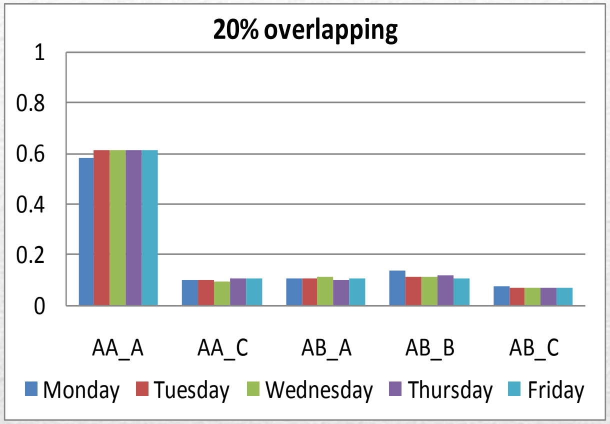
Goodness of Fit Considering Overlaps

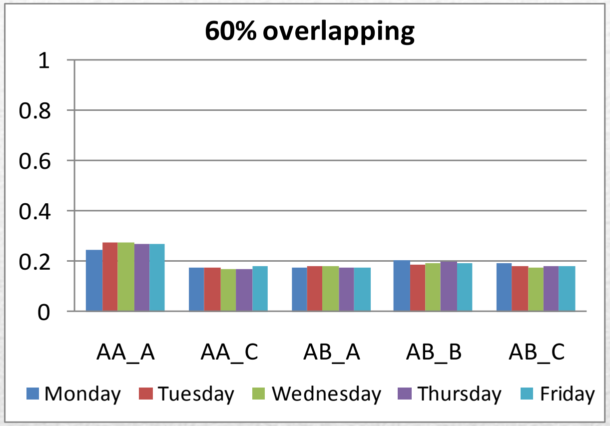


- The likelihood ratio index (ρ) improved significantly for very low overlapping thresholds S (i.e. 20% is better than 40% and so on).
- No improvement, compared to ignoring overlapping, can be observed for S>40%.
- Considering overlap is important to increase the fitness of the predicted model.

Discussion







- Only for S <40%, the variation in route choice decreases compared to route choice prediction without considering overlapping.
- A large part of the route choice variation might indeed be due to common lines that are part of the travelers' hyperpaths or at least some route variation is due to overlap and possibly hyperpaths.

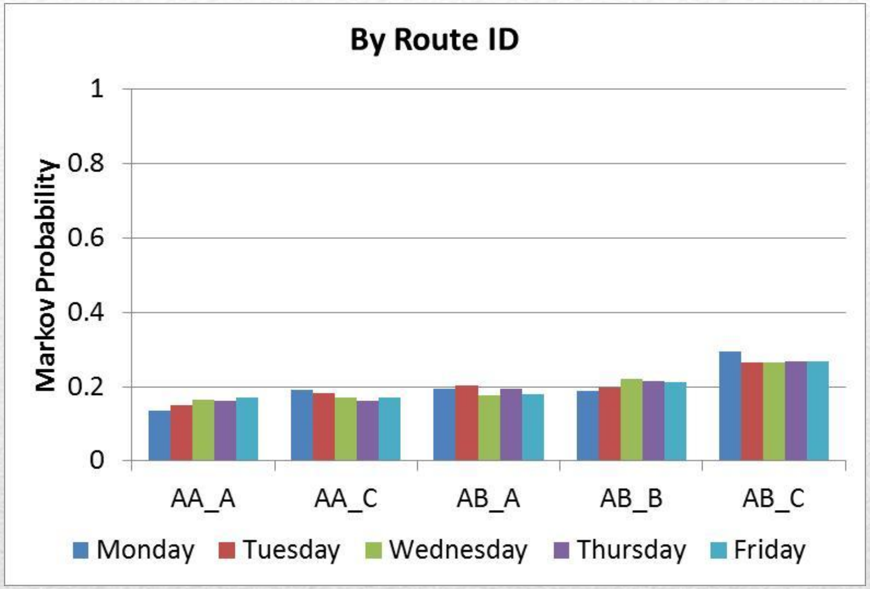
Comparison of Route Choice Variation and Bus Stop Choice Variation

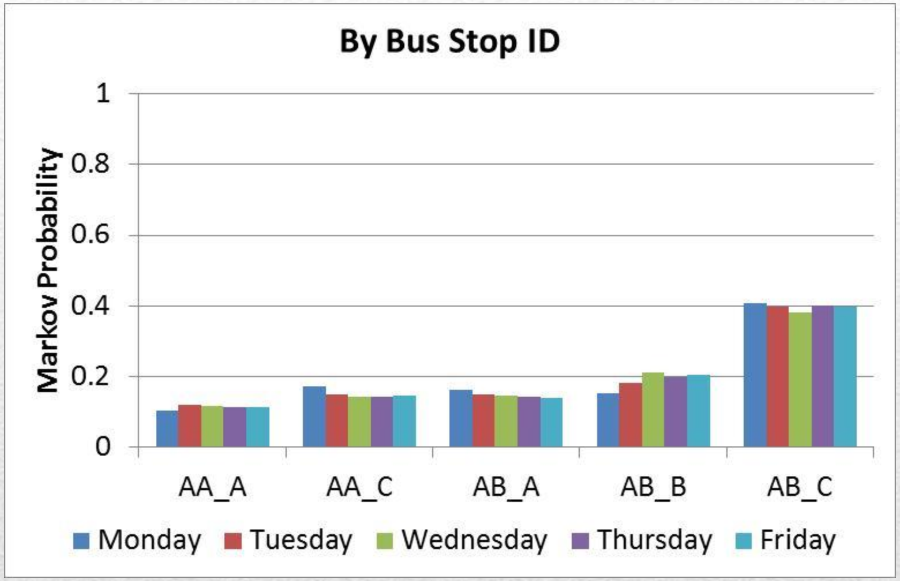
- New three weeks Oyster Card data (16 Aug ~ 5 Nov, 2011) was obtained
  - More reliable boarding location data as well as boarding route, boarding time and card ID data are available
- Comparing passengers' route choice variation and boarding bus stop choice variation

Hypotheses : *passengers would travel on hyperpaths if their route choice variation is larger than their bus stop choice variation*

- Data extraction
  - As well as previous analysis, picking up the traveller who use bus on all 10 working days (the 1st week of the data and the 2nd week of the data, or the 2nd week of the data and the 3rd week of the data) before 9.30 am

Discussion





- The Markov probability of AA_A by route ID is larger than that by bus stop ID and the Markov probability of AB_C by route ID is smaller than that by bus stop ID.
- Contrary to our expectation, the variation of route choice by route ID is smaller than that by bus stop ID.
- Passengers may shift the boarding stops…

Conclusions

- Smart card data might be good to understand traveller behavior patterns, but has a lot of limitation.
- Markov models can be used to analyse the consistency in travelers' route choice behavior.
- The day of the week for which the route choice is predicted does not appear to have an influence on the results.
- There is some random variation in routes chosen, but not very sure if this is an outcome of hyperpath choice, or stochastic choice with uncertainty.

Further Trials

- Different data extraction strategy
  - Using same station every day!?
- Looking at different time of day (with more flexibility)

- Trial on other smartcard data
  - We have been analysing AYUCA card data in Gifu.
    - 8 week data (Oct, Nov, 2011) with 82,320 travellers /2,100,285 trips,
    - Boarding/alighting stops recorded,
    - Bus is the main public transport mode in Gifu.
    - Schedule-based service.