WHERE ARE THE PASSENGERS? INDIVIDUAL PASSENGER ROUTE CHOICE ESTIMATION FROM GATE INFORMATION IN URBAN RAIL TRANSIT SYSTEMS

Thomas James Tiam-Lee and Rui Henriques
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa

1. INTRODUCTION

Public urban rail transit is a major mode of transportation in many cities around the world. The latest report by the International Association of Public Transport (2018) estimates that metro systems had a total global ridership of 53768 million in 2017, with Asia-Pacific and Europe leading the numbers. The massive usage of urban rail transit underscores the importance of government efforts to ensure that metro systems are reliable, safe, and efficient for the public. Among the challenges tackled by metro operators is the effective resolution of bottlenecks in passenger traffic, including insufficient capacity for passenger demand, a task that has even more relevance in the context of the current COVID-19 pandemic to ensure the satisfaction of health safety norms.

Passenger route choices are not deterministic as they depend on the subjective perception of travel time, required transfers, convenience factors, and on-site train arrivals and waiting times, among others (Daaman, Bovy and Hoogendoorn, 2005; Xinyue, Liping, Haiying et al., 2018). This makes it difficult to infer the volume of passengers along specific segments of the network at a given time.

In this study, we model individual passenger routes and the overall passenger flow in a rail network based only on automated fare collection data containing passenger entrances and exits within the urban rail system. To accomplish this task, we present a computational approach that assesses the likelihood of each possible route choice by aligning card validation timestamps against real-time route scheduling. In the absence of vehicle geolocation data, the locations of the trains at different times can be estimated by analyzing passenger volume peaks at the exit station gates. We can then use this information together with an analysis of the trip durations to better infer the likelihood that a passenger took a specific route by analyzing the location and timings of his or her entry and exit from the urban rail system.

The approach offers two main advantages. It requires minimal information that is easily obtainable in most urban rail transit systems. This allows it to be implemented easily without the use of complex and privacy-invasive sensors and monitoring devices. Previous approaches require additional information such as models representing initial passenger route choice behaviors (Xu, Xie, Li, et al., 2018) and train arrival times (Sun & Xu, 2012; Zhao, Zhang, Tu, et al., 2016). Second, it is adaptable to changes in the network not only because of unforeseen events like malfunctions and delays, but also changes in operational schedules and policies. This makes it robust and reliable in the presence of changes over time unlike previous approaches that assume fixed and precise train arrival times for each station (Sun & Xu, 2012; Zhao, Zhang, Tu, et al., 2016).

In this paper, we apply the said approach on the automated fare collection dataset of the Lisbon Metro for the month of October 2019 and discuss the corresponding results. The paper is structured as follows. Section 2 contains a brief background of the Lisbon Metro and the automated fare collection dataset used in this paper. Section 3 introduces the proposed approach for estimating passenger route choices and passenger flow. Section 4 shows the results and validation of the said approach, and Section 5 contains the concluding remarks.

2. BACKGROUND

2.1 Description of the Task

The targeted task is to estimate the route choices of passengers in a mass transit system, in which the system is composed of a set of interconnected lines, with each line serviced periodically by vehicles passing through a series of stations. Passengers may enter and exit the system through any station and may take as vehicles as many times as they want while within the network. We define the route of a passenger as the path taken from the station of entry to the station of exit. We then formalize the task as follows: given the passengers' time and location of entry and exit in the transit system, how can we predict the actual routes that were taken by each passenger?

2.2 Lisbon Metro

In this paper, we use the Lisbon Metro as an example application domain for the passenger route choice estimation approach. The Lisbon Metro is an urban rail rapid transit system in Lisbon, the capital city of Portugal. As of 2021, it consists of 4 lines: *azul* (blue), *amarela* (yellow), *verde* (green), and *vermelha* (red) with a total of 56 stations (Metropolitano de Lisboa, E.P.E., 2021). As with most other urban rail transit systems, the different lines interconnect with one another through certain common stations. Figure 1 shows a map of the rail network.

As the rapid transit system of a capital city, the Lisbon Metro serves a massive number of passengers. In 2019, the annual ridership of the Lisbon Metro reached 173 million people (Miguel, 2020). Similar with most other urban rail systems, the Lisbon Metro makes use of an automated fare collection system, in which passengers use cards to enter and exit the stations. After entering, the passenger is free to ride any train within the network as many times as they want, before exiting at any station. With each entrance and exit made by each passenger, the automated fare collection system records the following information: the timestamp (date and time), the station, and a unique identifier for the passenger (i.e., the number of the card used for the entry or exit). The collection of all such records from the automated fare collection system forms the dataset that we will be using for estimating passenger flow and route choices within the network. Specifically, we will be focusing on data for the month of October 2019 containing over 33 million entrances and exits.

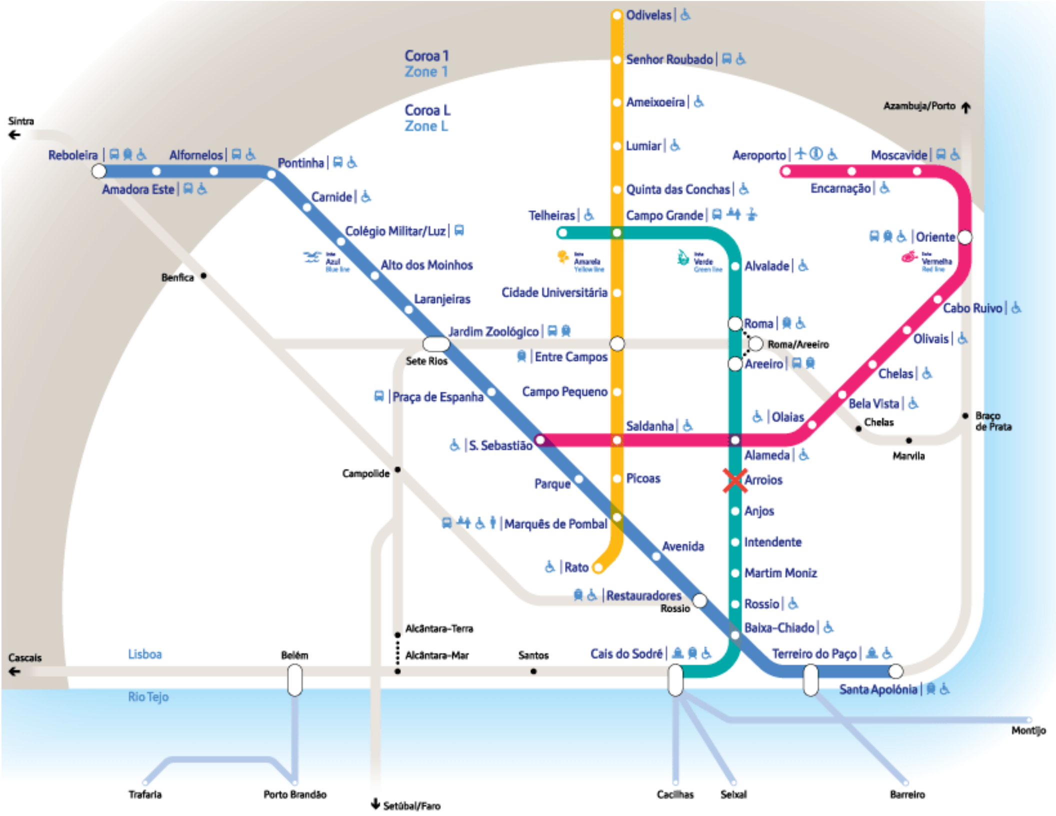

Figure 1. Lisbon Metro network map.

3. ESTIMATION OF PASSENGER ROUTE CHOICES

In this section, we present a two-step approach for estimating passenger route choices in urban rail transit systems. The approach is divided into two parts: (a) estimating the locations of the trains in the network, and (b) estimating the likelihood of each passenger's route choice based on the train locations.

3.1 Estimation of Train Locations

The first step is to estimate the locations of each train in the network over time. To this end, peaks in the volume of passengers exiting the stations are identified based on automated fare collection data. For any given day, we can form a time series for each station by aggregating the number of exit records per station into 15-second intervals. We then perform a rolling average and rolling deviation on the time series to identify passenger exit peaks along a line to estimate the precise location of vehicles. We define a passenger volume spike as a peak in the time series that is at least one standard deviation higher than rolling mean. This scheme aims to ensure that only peaks deviating from typical behavior along a given period are considered and has previously been used similarly for robust peak detection in other studies (Dons, Laeremans, Orjuela et al., 2019; Cloud, Tarlen, Liu et al., 2019).

Using the Lisbon Metro as an example, Figure 2 shows the aggregated passenger exits on Santa Apolónia station on a single day. The red dots represent the passenger volume spikes, as identified through the peaks that are at least one standard deviation away from the rolling mean. From this, it is apparent that there are periodic spikes in the number of people exiting the station, caused by groups of people alighting from the trains that arrive periodically. We identify such spikes in passenger volume and use them as a basis for estimating the train locations within the network.

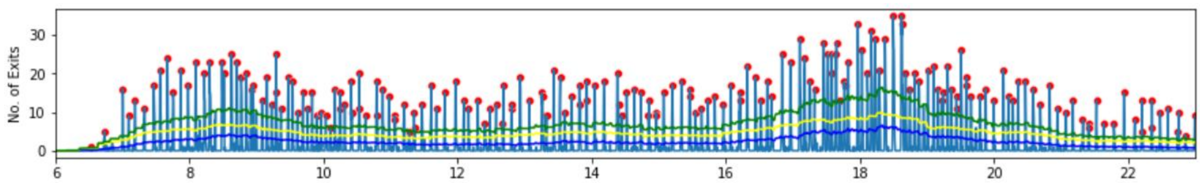

Figure 2. Volume of passengers over time for October 7 in Santa Apolónia station. Red dots show spikes in passenger volume, blue line shows the rolling mean, yellow line shows the rolling standard deviation, and green line shows one standard deviation away from the rolling mean.

We then identify the travel time durations of each train through each line. This data can easily be obtained based on the urban rail system's operational protocols or, in the absence of this data, could be estimated from averaging times of actual train operations. Given a line consisting of stations $s_1, s_2, \dots, s_n$, the train arrival offset at station s_i, referred to as $\mathrm{offset}(s_i)$, is the total amount of time it takes for a train to arrive at a station s_i from station s_1. Thus, $\mathrm{offset}(s_i)$ can be defined recursively, where $\mathrm{offset}(s_1) = 0$ and $\mathrm{offset}(s_i) = \mathrm{offset}(s_{i-1}) + \delta(s_{i-1}, s_i)$, where $\delta(s_{i-1}, s_i)$ is the time takes for the train to arrive to s_i from s_{i-1}. Note that the train arrival's offsets on the same line moderately differ from the reverse direction.

Using the Lisbon Metro's green line as an example, Figure 3 shows the train arrival offsets of each station for both directions.

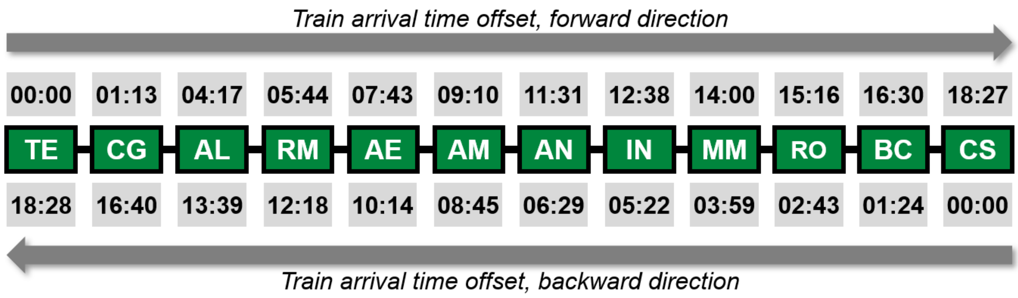

Figure 3. Train arrival time offsets from the first station in *Linha Verde* (green line).

The next step is to align the passenger volume peak data for each station according to the projected passing of the train along the stations using the previously identified offsets. To do this, we perform a shift by subtracting the corresponding train arrival offset from each of the passenger volume peaks in each station. We then estimate the likelihood that a train started on the first station at a given time. To this end, we define a likelihood score based on the total distance from the closest passenger volume spikes. Given a starting time t, the likelihood score $\mathcal{L}(t)$ is computed as:

$$\mathcal{L}(t) = 1 - \frac{1}{\alpha^2} \sum_{s \in S} \frac{(closest(s,t,\alpha) - t)^2}{count(s,t,\alpha) + 1}$$

where S is the set of stations in the target line and α is a constant time duration threshold. The $closest$ function returns the time of the closest passenger volume peak (after shifting) that occurs after t, bounded by a maximum value of $t + \alpha$. The $count$ function returns the total number of exits made within the range $[t - \alpha, t + \alpha]$ (after shifting). Intuitively, the numerator represents the squared error of the train arrival times for each station from the closest passenger volume peak, while the denominator weighs the errors according to the volume of passengers to give more importance to stations with larger volumes of passengers around that time. The 1 constant is added to ensure that the operation is defined even in the absence of passenger exits. Finally, the summation of the errors is normalized to the range $[0,1]$, where 1 represents a perfect alignment of the train arrivals with passenger volume peaks. We can identify the times where trains are likely to have started from the first station of the line by identifying the peaks of the likelihood score plot.

Illustrating, Figure 4 shows the times of the passenger volume spikes in *Linha Verde* (green line) before and after shifting. Based on our assumption that the arrival of a train on a station implies a possible spike in passenger volume at the exit gates, we are able to draw a vertical line on the shifted plot and observe volume spikes close to that line. We can visually observe in the figure that this hypothesis holds true, particularly on the latter half of the line. This makes sense because people are more likely to get off at later stations according to the natural flow of the line direction. Figure 5 shows $\mathcal{L}(t)$ computed for various values of t, showing the likelihood peaks at times where the passenger volume peaks are well-aligned, with the purple dots showing the peaks which represent the predicted times that the trains have started travelling from the first station in this direction.

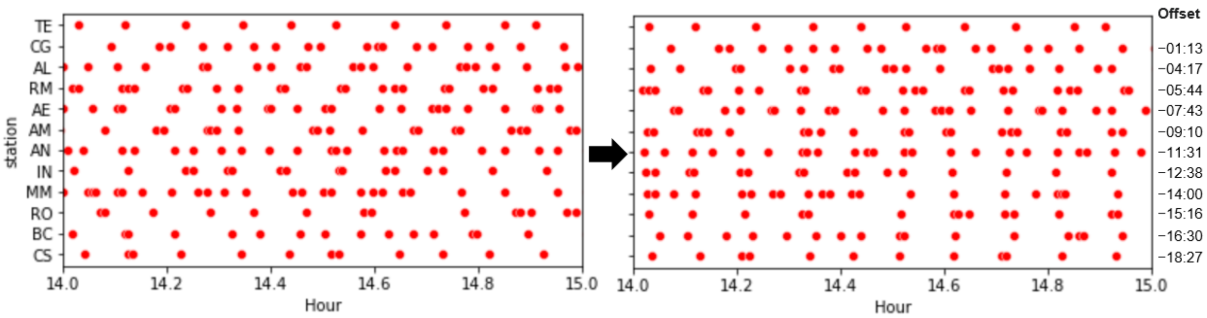

Figure 4. Passenger volumes spikes per station on *Linha Verde* (green line) before and after shifting according to train arrival time offsets.

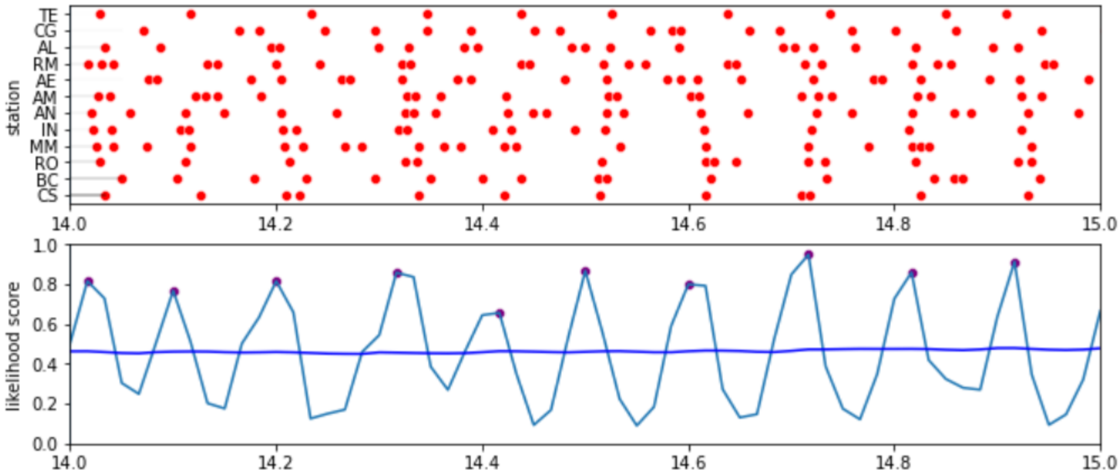

Figure 5. Likelihood score computed across different times, showing the score peak at times where the passenger volume peaks are well-aligned.

Now that the predicted starting times of the trains have been identified, actual train arrival times for each station can be easily inferred. To compute the train arrival times at a given station s_i, we simply add the $\text{offset}(t_i)$ to each of the estimated starting times from the lines passing through s_i. Using this information, given a passenger's entrance and exit information from the automated fare collection data, potential routes taken from entrance to exit can now be explored.

3.2 Prediction of Passenger Route Choices

Before route choices are predicted, we first pair the entrance and exit records in the automated fare collection dataset. Given that each record contains the following information: (a) entrance or exit, (b) timestamp, (c) station, and (d) identifier, we can match each passenger's entrance record with its corresponding exit record considering the card identifiers and trip precedencies. There may be some cases where an entrance or exit record is missing its corresponding pair in the dataset. These can occur due to operational anomalies (e.g., malfunctions, cases where passengers were allowed to enter/exit without passing through the gates in extraordinary situations). Understandably, route choice estimation for incomplete trips should be proceeded by well-established principles for boarding or alighting station inference (Cerqueira et al. 2020). Nevertheless, the likelihood of incomplete trip records for rail systems with closed gates is generally low.

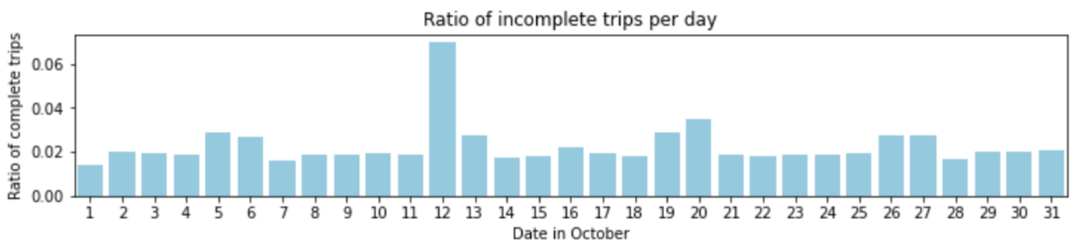

Figure 6. Ratio of incomplete trips over all trips per day in the Lisbon Metro.

In fact, after pairing the entrance and exit records in the Lisbon Metro, Figure 6 shows the ratio of incomplete trips (i.e., missing entrance or missing exit) over all trips for each day in the month of October. Overall, only 2.09% of the trips are incomplete, although there was an unusually high number of incomplete trips on the 12th of the month, likely caused by an unusual operational event. To remove the additional uncertainty associated with route choices along incomplete trip records, we excluded these records from the conducted experimental analysis.

We define a route as a path that a passenger takes to get from an entry to an exit station. A route may contain one or more segments, which represent a single train ride on any given line. Given a passenger's time and location of entry and a candidate route, the expected exit time of the passenger by following that route can be computed. To estimate the expected exit time along a route, travel can be simulated considering the train arrival times for the relevant stations.

To accommodate for cases where a passenger misses the next available train (e.g., train is full, walked too slowly from gate to the train platform, waiting for a friend), we introduce a parameter max_lag which represents the maximum number of trains skipped per segment of the route. Thus, for each route the algorithm returns a set of expected exit times, each computed based on the number of trains skipped in different segments of the route. To formalize, given the time of entrance and a candidate route, we estimate the time of exit using the following algorithm:

```
curr_time := {time of entrance}
foreach segment in route:
     source, dest := start and end station of the segment
     train_arrivals := set of arrival times of trains at source
     travel_time := offset(dest) - offset(source)
     next_time := {}
     foreach curr in curr_time:
          next_train_idx := index of the next train arrival from curr
          for i in 0..max_lag:
               next_time.push(train_arrivals[next_train_idx + i] + travel_time
     curr_time := next_time
return curr_time
```

Given a set of estimated exit times $R = \{r_1, r_2, \dots\}$ from a specified route and the actual exit time t_{exit} of the passenger, an error score can be computed as $\min_{r_i \in R}(r_i - t_{exit})^2$. The error scores of different candidate routes could be compared to assess the likelihood that each route was taken, with a smaller score representing a better likelihood. A threshold can also be placed to ensure that the error is small enough to assert confidence in the prediction.

In the Lisbon Metro dataset, consider for instance a passenger who entered through Alameda station at 12:59:15 and exited through Campo Grande station at 13:11:27. Two possible routes that this passenger could have taken are as follows: (a) green line from Alameda to Campo Grande and (b), red line from Alameda to Saldanha, followed by the yellow line from Saldanha to Campo Grande. As visualized in Figure 7, from the time of entry, the expected exit time if route (a) was taken is 13:10:40, while the expected exit time if route (b) was taken is 13:22:56. Since the actual duration of the passenger's trip was 13:11:27, it is more likely that route (a) was taken.

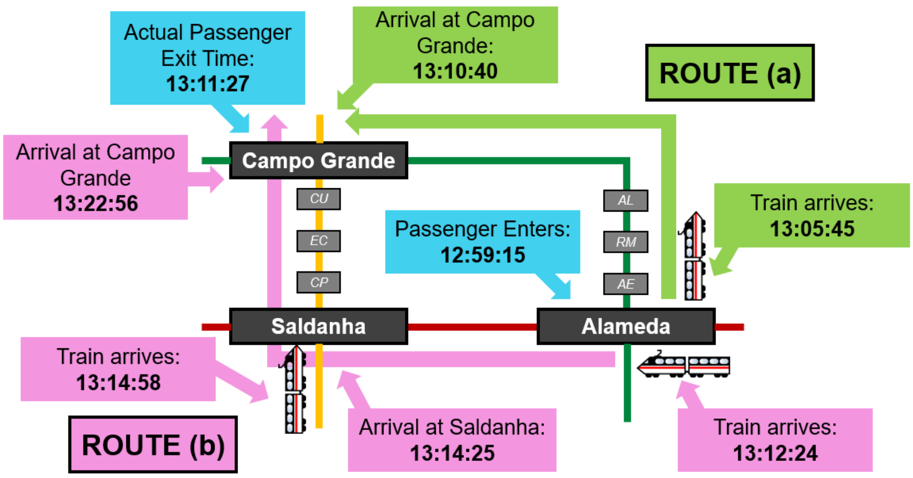

Figure 7. Two possible routes from Alameda to Campo Grande, shown in light green (a) and pink (b). It is more likely that the passenger took route (a) based on the distance between the expected arrival at Campo Grande and the actual exit time.

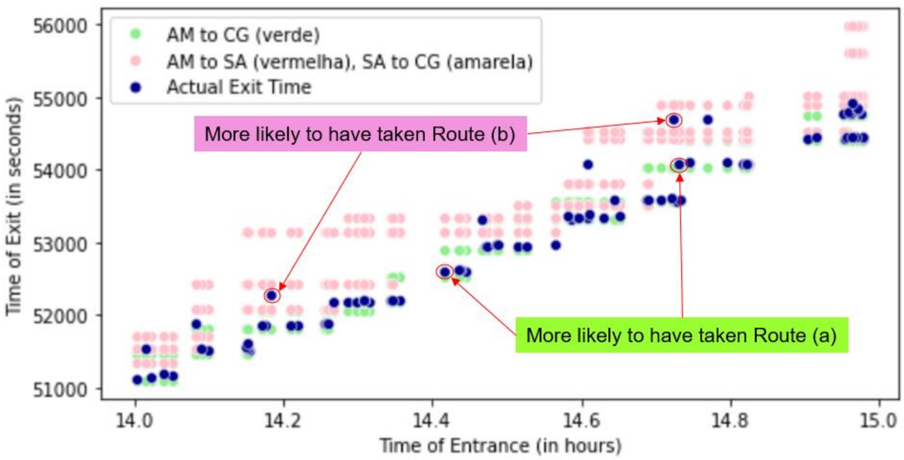

Figure 8. Actual passenger exit times in comparison with the expected exit times for two different routes from Alameda to Campo Grande.

In Figure 8, we can see a plot of the actual exit times of each passenger who entered through Alameda and exited through Campo Grande, as well as the expected exit times of route (a) in light green and route (b) in pink, both computed with a max_lag

of 1. Most passenger exits are closer to the expected exit times of route (a), which is the more direct path that does not require any transfers. Nevertheless, it can be seen that there are a few passengers likely to have taken the other route based on their exit timing.

4. RESULTS

We apply the proposed algorithm to the Lisbon Metro dataset along October, 2019, with the analyses being incident on October 7th. First, we identify all pairs of stations within the network. For each pair $(s_a, s_b), s_a \neq s_b$, we identify all candidate route choices from station s_a to s_b though a standard exhaustive graph search algorithm. To limit the candidate route choices, we impose the restriction that a route cannot contain the same station twice. We then apply the proposed approach to predict the route that is more likely to have been taken by each passenger who entered through s_a and exited through s_b.

We used a max_lag of 1, based on the assumption that most passengers will not skip trains apart from cases passenger did not walk fast enough to the platform to catch the next available train. We consider $\alpha = 180$ seconds, which is chosen as a time window wide enough to capture a single episode of a passenger volume spike in a station. We also used a minimum error threshold of $180^2 = 32400$ to only consider predictions that are within 180 seconds of the closest expected exit time. Because of this, not all passenger trips are assigned a predicted route choice, i.e., if the exit time is not within 180 seconds of the expected exit times of the candidate routes, the route choice is deemed “unknown”. This is to be expected given the presence of small markets, stores and coffee shops within the network. In the dataset, 89% of the trips were successfully assigned to a candidate route choice (i.e., not “unknown”).

In our analysis, we highlight two common criteria for passenger preference: (a) least number of transfers between lines (a measure of convenience and time), and (b) least number of crossed stations (a measure of distance and time). There are a total of 49 stations across all lines in the metro, resulting in 2352 possible entrance-exit pairs. Out of these pairs, 636 (27.04%) are non-trivial entrance-exit pairs where the route with the least transfers is different from the route with the least number of stations. On the other, 1716 (72.95%) are entrance-exit pairs where there exists an ideal route that has **both** the least number of transfers and the least number of stations. We refer to these pairs as trivial pairs since the passenger’s route choice in these cases is straightforwardly met under the above criteria.

Among the 636 non-trivial station pairs, the preferred route is the one with least transfers in 562 (88.36%) of the pairs. On the other hand, the preferred route is the one with the least number of stations in 54 (8.49%) of the pairs. As for the remaining pairs, either the preferred route was neither of the two, or it could not be determined because no trips were made under that pair for that day. Figure 9 shows the distribution of the ratio of passengers who took the route with the least transfers and the route with the least stations across all non-trivial pairs.

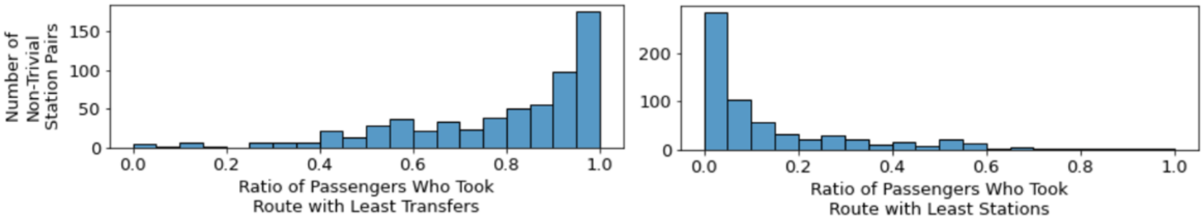

Figure 9. Ratio of passengers who took route with least transfers (left) and route with least stations (right) among non-trivial pairs.

Table 1. Some Examples Among Non-Trivial Pairs

	Entry Station	Exit Station	Route with Least Transfers	Ratio	Route with Least Distance	Ratio
(a)	AN	AS	AN → BC → AS (**1 transfer**, 18 stations)	**0.98**	AN → AM → SS → AS (2 transfers, **12 stations**)	0.02
(b)	TP	OR	TP → SS → OR (**1 transfer**, 14 stations)	**0.96**	TP → BC → AM → OR (2 transfers, **12 stations**)	0.02
(c)	CP	AE	CP → CG → AE (**1 transfer**, 6 stations)	**0.56**	CP → SA → AM → AE (2 transfers, **3 stations**)	0.44
(d)	AM	AV	AM → SS → AV (**1 transfer**, 5 stations)	**0.28**	AM → SA → MP → AV (2 transfers, **4 stations**)	0.01
			AM → BC → AV (**1 transfer**, 7 stations)	**0.7**		
(e)	PI	CS	PI → CG → CS (**1 transfer**, 15 stations)	0.11	PI → MP → BC → CS (2 transfers, **5 stations**)	**0.86**
(f)	CS	PI	CS → CG → PI (**1 transfer**, 15 stations)	0.06	CS → BC → MP → PI (2 transfers, **5 stations**)	**0.57**
(g)	JZ	IN	JZ → BC → IN (**1 transfer**, 10 stations)	0.47	JZ → SS → AM → IN (2 transfers, **6 stations**)	**0.51**

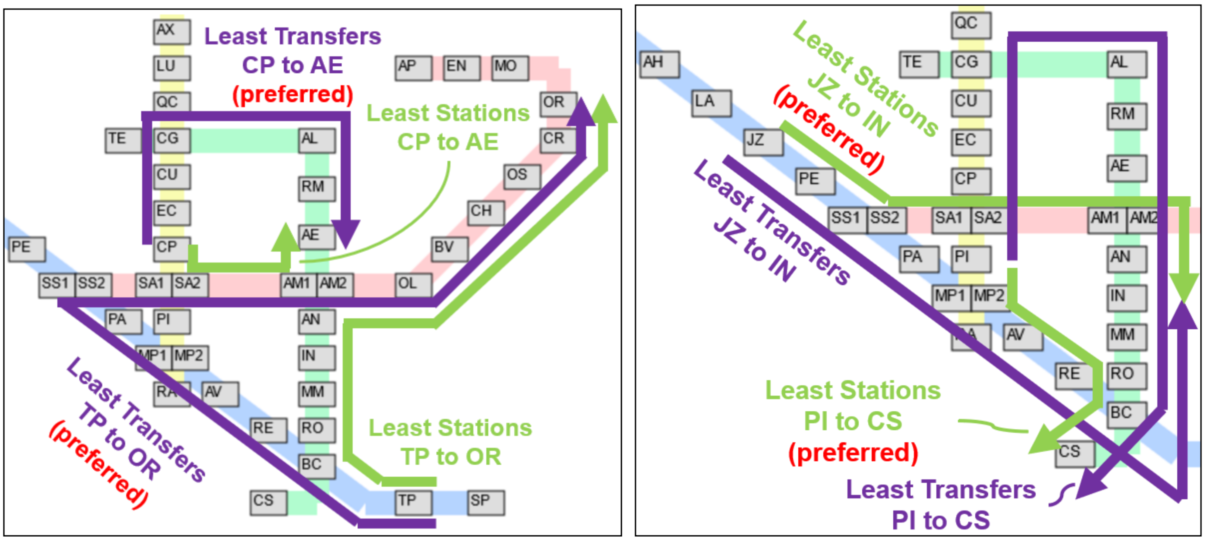

Figure 10. Visualizations of examples (b) and (c) where the route with least transfers is preferred (left) and examples (e) and (g) where the route with least stations is preferred (right).

Table 1 shows selected cases among non-trivial pairs. In examples (a) to (d), the route with the least transfers was the preferred route, while in examples (e) to (g) the preference is towards routes with the least number of stations. As shown in Figure 10, for passengers travelling from Campo Pequeno (CP) to Areeiro (AE), more passengers preferred to take the route with only one transfer, even it had a longer distance. However, this is not always the case, as for passengers travelling from Jardim Zoologico (JZ) to Intendente (IN), slightly more passengers took the route with more transfers but with lower distance. This is likely because of the perceived trade-off between convenience and speed. In the cases where people chose the shorter route over the one with the least transfers, the latter route appears to be significantly longer.

Among the trivial pairs, on the other hand, Figure 11 shows the ratio of passengers choosing the ideal route (least transfers *and* number of stations). Unsurprisingly, the ideal candidate route is the preferred route choice in almost all cases. However, there are cases where evidence suggests this is not the case, potentially caused by card profiles who undertake recreative or working activities along the network. Table 2 shows some pairs that show this.

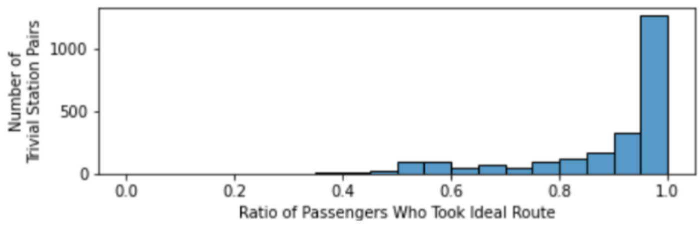

Figure 11 Ratio of passengers who took ideal route in trivial station pairs.

Table 2. Some Cases Where Many Preferred an Alternative to the Ideal Route Among Trivial Pairs

	Entry Station	Exit Station	Ideal Route	Ratio	Alternative Route	Ratio
(h)	PO	SA	PO → SS → SA (**1 transfer**, **8 stations**)	0.4	PO → MP → SA (1 transfer, 11 stations)	0.58
(i)	AH	SA	AH → SS → SA (**1 transfer**, **5 stations**)	0.4	AH → MP → SA (1 transfer, 8 stations)	0.59
(j)	CM	SA	CM → SS → SA (**1 transfer**, **6 stations**)	0.44	AS → MP → SA (1 transfer, 9 stations)	0.53
(k)	AP	BC	AP → AM → BC (**1 transfer**, **14 stations**)	0.36	AP → SS → BC (1 transfer, 16 stations)	0.55
(l)	EN	BC	EN → AM → BC (**1 transfer**, **13 stations**)	0.44	EN → SS → BC (1 transfer, 15 stations)	0.52

In examples (h), (i), and (j), it appears that under the prediction parameters, although the ideal route to Saldanha (SA) from various stations in the northern half of the blue line is through a transfer in São Sebastião (SS), relatively higher number of individuals preferred to transfer in Marquês de Pombal (MP) instead. This observation can be partially explained by the large walking distance associated with the Alameda's transfer. Similarly, in examples (k) and (l), a comparable number of passengers chose to transfer in São Sebastião (SS) instead of Alameda (AM) on the way to Baixa-Chiado (BC). These examples demonstrate instances where the predicted route choice contradicts the obvious choice and may be of interest to metro administrators.

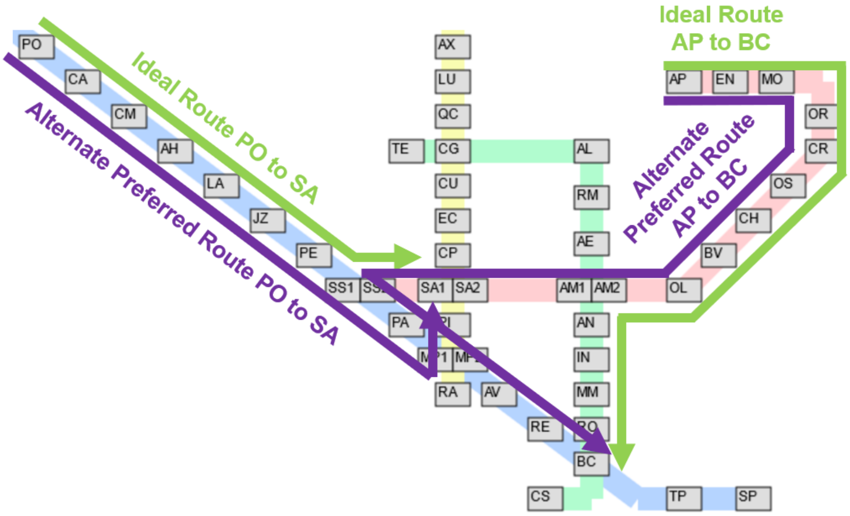

Figure 12 Visualizations of selected examples (h) and (k).

5. CONCLUSION

In this paper, we proposed a novel approach for predicting passenger route choices in a transport network using automated fare collection data. In contrast with state-of-the-art principles on route estimation, the proposed method requires minimal card ticketing information and is parameterizable to contemplate the missing of trains due to arbitrarily-high movement from gate to the platform and unavailable capacity at vehicles at peak hours. The proposed method is further robust to unforeseen events, such as malfunctions and operational delays, and dynamically adaptable to topological changes in the network as long as those are standardly captured in the reference General Transit Feed Specification source.

We applied the approach to a real trip record data collected from the Lisbon metropolitan system and found that different users place different choices along the same entry-exit station pairs. Although the majority of preferred routes along the network corresponds to those with the least transfers, a considerable amount of choices do not follow this assumption. In our manuscript, we comprehensively present cases wherein the route's choice by the majority of passengers is guided by the shortest distance and not by the number of transfers.

The implications of these observations are paramount, highlighting the need for an accurate estimation of route choices as they can considerably change the current perceptions on line demand and in-vehicle occupation levels. In this context, the need for more robust principles, as the ones placed in this work, are essential for detecting demand-driven bottlenecks, particularly relevant to ensure the satisfaction of safety norms during pandemic periods. The fine grained location of patients along a urban rail transit systems can therefore support the vehicle scheduling and their re-capacitation along different periods, as well as guide complementary initiatives for the positive conditioning of route choices.

ACKNOWLEDGMENTS
The authors thank METRO and Câmara Municipal de Lisboa (Gabinete de Mobilidade and Centro de Operações Integrado) for the provision of data and support. This work was further supported by national funds through Fundação para a Ciência e Tecnologia (FCT) under projects WISDOM (DSAIPA/AI/0044/2018) and ILU (DSAIPA/DS/0111/2018), and INESC-ID pluriannual (UIDB/50021/2020).

BIBLIOGRAPHY

Cerqueira S., Henriques R. & Arsénio E., Integrative Analysis of Traffic and Situational Context Data to Support Urban Mobility Planning, European Transport Conference (ETC), 2020

Cloud B., Tarlen B., Liu A. et al., Adaptive Smartphone-Based Sensor Fusion for Estimating Competitive Rowing Kinematic Metrics, San Francisco, PloS one 14.12, 2019.

Daaman W., Bovy P., Hoogendoorn S., Influence of Changes in Level on Passenger Route Choice in Railway Stations, Newbury Park, Transportation Research Record, 2005.

Dons E., Laeremans M., Orjuela J. et al., Transport Most Likely to Cause Air Pollution Peak Exposures in Everyday Life: Evidence from Over 2000 Days of Personal Monitoring, Amsterdam, Atmospheric Environment 213, 2019.

Inês M., Metro de Lisboa tem Novo Ano Recorde. Empresa Transportou 173 Milhões de Passageiros em 2019, Retrieved from https://jornaleconomico.sapo.pt/noticias/metro-de-lisboa-tem-novo-ano-recorde-empresa-transportou-173-milhoes-de-passageiros-em-2019-540746, 2020.

Metropolitano de Lisboa, E.P.E., Metro Lisboa, Retrieved from https://www.metrolisboa.pt/en/, 2021.

Sun Y., Xu R., Rail Transit Travel Time Reliability and Estimation of Passenger Route Choice Behavior: Analysis Using Automatic Fare Collection Data, Newbury Park, Transportation Research Record, 2012.

World Metro Figures, Statistics Brief, Brussels, International Association of Public Transport, 2018.

Xu X., Xie L., Li H. et al., Learning the Route Choice Behavior of Subway Passengers from AFC Data, Amsterdam, Expert Systems with Applications, 2018.

Zhao J., Zhang F., Tu, L. et al., Estimation of Passenger Route Choice Pattern Using Smart Card Data for Complex Metro Systems, IEEE Transactions on Intelligent Transportation Systems, 2016.