EveryoneCounts: Data-Driven Digital Advertising with Uncertain Demand Model in Metro Networks

Desheng Zhang [§*], Ruobing Jiang[†*], Shuai Wang[§], Yanmin Zhu[†], Bo Yang[†], Jian Cao[†], Fan Zhang[‡], Tian He[†§*]
[§]University of Minnesota, USA, [†]Shanghai JiaoTong University, China [‡]SIAT, China

***Abstract*—Nowadays most metro advertising systems schedule advertising slots on digital advertising screens to achieve the maximum exposure to passengers by exploring passenger demand models. However, our empirical results show that these passenger demand models experience uncertainty at fine temporal granularity (e.g., per min). As a result, for fine-grained advertisements (shorter than one minute), a scheduling based on these demand models cannot achieve the maximum advertisement exposure. To address this issue, we propose an online advertising approach, called *EveryoneCounts*, based on an uncertain passenger demand model. It combines coarse-grained statistical demand modeling and fine-grained bayesian demand modeling by leveraging real-time card-swiping records along with both passenger mobility patterns and travel periods within metro systems. Based on this uncertain demand model, it schedules advertising time online based on robust receding horizon control to maximize the advertisement exposure. We evaluate the proposed approach based on an one-month sample from our 530 GB real-world metro fare dataset with 16 million cards. The results show that our approach provides a 61.5% lower traffic prediction error and 20% improvement on advertising efficiency on average.**

I. Introduction

Digital advertising systems [1] in metro networks obtain increasing preference by advertisers because of the extensive exposure to a large number of passengers, e.g., in New York City, the average ridership reaches up to 5.5 million [2] according to the Metropolitan Transportation Authority. In such a metro advertising system, the key design issue is how to schedule given advertising time (rented by advertisers) on a digital screen to maximize passing audience for the maximum advertisement exposure during all the advertising time, which is a key indicator of the potential advertising revenues [1].

Historically, an intuitive approach for this issue is to preferentially display advertisements in those time periods with more historical passengers, i.e., a passenger demand model based on historical passenger data collected in the automatic fare collection (AFC) system [3]. The assumption behind such an approach with a historical-data-based demand model is that the daily distribution of passenger traffic volumes, i.e., passenger demand, in the same metro station is fairly predictable with historical distributions. Thus, higher historical passenger demand in a past time period indicates higher future passenger demand for the same time period in the future day.

However, our empirical study in the metro network of Shenzhen (the twin city of Hong Kong and has more than 13 million residents) clearly shows that only coarse-grained (e.g., one hour) passenger demand is predictable based on historical data, but fine-grained (e.g., one minute) passenger demand is with uncertainty and thus unpredictable based on historical data. Further, we argue that the fine-grained (instead of coarse-grained) passenger demand model is essential to advertising time scheduling, because an advertisement is normally in a fine-grained length (e.g., less than one minute). As a result, we face a key challenge to predict fine-grained passenger demand to effectively schedule adverting time.

To address this challenge, we propose an online advertisement scheduling approach, named *EveryoneCounts*, based on robust receding horizon control with an uncertain demand model. In EveryoneCounts, it combines coarse-grained statistical demand modeling and fine-grained bayesian demand modeling by inferring individual arrivals using both real-time and historical, *instead of only historical*, data in the AFC system. This is because given real-time entering stations and time, we can accurately infer destinations and arriving time for passengers based on the low conditional entropy of passenger destinations and predictable travel durations in the metro network, learned from our extensive empirical study. In addition, we design an adverting scheduling algorithm based on receding horizon control to maximize advertisement exposure under our uncertain demand model. Specifically, our main contributions are as follows:

- To our knowledge, we perform the first work, named EveryoneCounts, to optimize advertisements exposure with large-scale data from the metro automatic fare collection system. We will release our data for the benefit of research community after the paper is accepted.
- We design a two-level uncertain metro passenger demand model based on our empirical study, which reveals that the coarse-grained passenger demand is predictable with the historical data while the fine-grained passenger demand is with uncertainty and thus unpredictable based on history. Thus, we combine coarse-grained statistical demand modeling and fine-grained bayesian demand modeling by leveraging (i) the real-time card swiping records collected by the AFC system, (ii) the low conditional entropy of passenger destinations, and (iii) the predictable travel time between different stations.
- We propose an online digital advertising approach based on robust receding horizon control, which exploits our passenger model to allocate coarse-grained advertising time offline but adjust fine-grained advertising time online to maximize advertisement exposure.
- We evaluate the performance of our approach through

[*]As co-first authors, the first two authors have equal contribution, and Tian He is the corresponding author.

extensive trace-driven evaluation based on one-month sample from our 530 GB real-world metro fare dataset with 16 million cards in Shenzhen, China. Compared to statistical approaches, the proposed approach has a 61.5% lower prediction error of fine-grained traffic volumes, leading to a 20% improvement in advertising efficiency.

The rest of the paper is organized as follows. Section II defines the advertising optimization. Novel empirical results are presented in Section III. Secions IV- VII show the detailed design and its performance. Secion VIII discusses the real-world issues. We introduce the related work in Section IX and conclude the paper in Section X.

II. Models and Problem Definition

In this section, we first present the scenarios and the advertising models. Then, we provide the formal definition of the advertising efficiency optimization problem. The mainly used notations are summarized in Table I.

A. Scenarios and Models

We focus on the metro stations where passengers need to swipe their metro cards at the entrance/exit to enter/leave a station. We model the metro network as a set of connected metro stations, denoted by $\Psi = \{\psi_1, \cdots, \psi_d, \cdots, \psi_m\}$. As shown in Fig. 1, we show the metro network of Shenzhen and a scenario where digital screens are installed and exposed to passengers between AFC machines and platforms of a station. The lighter the icon, the higher the average passenger demand.

Figure 1. An example of the digital screen advertising in Shenzhen metro network: all the passengers from other stations will pass by the target screen when they leave their destination station.

The metro network sells advertising time of digital screens in all stations (assuming one screen per station) to advertisers and charges them based on the length of advertising time. The maximum length of daily advertising time for the digital screen at all stations, denoted by T (e.g., 24 hours), is divided into small time slots with length τ (e.g., one minute). Suppose γ advertisers buy advertising time on digital screens and the lengths of their advertising time are α_d^i, where $1 \leq i \leq \gamma$ and $1 \leq d \leq m$. We consider all advertisers with the same priority, although our method can also be used for multiple priorities.

TABLE I
Main Notations Used in This Paper

Notation	Description
T	The length of a time unit, e.g., 24 h, at a station ψ_d
τ	The length of time slots, e.g., 3 min
α_d^i	The length of advertising time for a station ψ_d
$\beta_{j\cdot d}^i$	The traffic volume of jth slot at a station ψ_d
$\lambda_{j\cdot d}^i$	The AD schedule for jth slot at a station ψ_d for advertiser i

For a given slot t_j and a station ψ_d, the passenger demand $\beta_{j\cdot d}$ is given by the *slot traffic volume*, i.e., the number of passengers passing by the digital screen during t_j when they exit/enter the station ψ_d.

B. Problem Definition

Given the advertisement utility (i.e., T) of metro stations, the objective of the metro advertising system is to provide the best advertising service to the advertisers in terms of optimized advertising efficiency. With the above models and notations, we now formally define the key term – advertising efficiency and its corresponding optimization problem.

Definition 1 (**Advertising Efficiency**). *The traffic volume over the total advertising time on the screens at all stations, i.e.,*

$$\frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}}\beta_{j\cdot d}\times\lambda_{j\cdot d}^i}{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\alpha_d^i}, \tag{1}$$

where the schedule of the jth slot at the station ψ_d for the advertiser i, $\lambda_{j\cdot d}^i$, is given by,

$$\lambda_{j\cdot d}^i = \begin{cases} 1, & \text{if } t_j \text{ is an AD slot for advertiser } i \text{ at station } \psi_d \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

For the numerator, the innermost summation is for all AD slots in a particular station ψ_d for a particular advertiser; the middle summation is for all AD slots in all stations for a particular advertiser; the outmost summation is for all AD slots in all stations for all advertisers. For the denominator, the summation is the total advertising time.

Definition 2 (**Advertising Efficiency Optimization**). *Given the total advertising time and the slot length, a set of schedules $\lambda_{j\cdot d}^i$ needs to be made, so that the advertising efficiency is maximized, i.e.,*

$$\max_{\lambda_{j\cdot d}^i} \frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}}\beta_{j\cdot d}\times\lambda_{j\cdot d}^i}{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\alpha_d^i},$$

$$s.t.\ \tau \times \sum_{j=1}^{\frac{T}{\tau}} \lambda_{j\cdot d}^i = \alpha_d^i, \forall i \in [1,\gamma], \forall d \in [1,m]. \tag{3}$$

To solve this optimization problem, we need to find a schedule $\lambda_{j\cdot d}^i$ for different AD slots in different stations for different

advertisers. In our setting, γ, m, T, τ, and α_d^i where $\forall i \in [1, \gamma]$ are given in advance. But the passenger demand $\beta_{j\cdot d}$ for a particular slot in a particular station has to be obtained by a demand model. Such passenger demand includes passengers who enter or exit the target station. But since entering passengers can be straightforwardly tracked by AFC machines at the entrance of the station, we focus on the exiting passengers who enter at other stations but exit at this station in this paper.

In this work, as shown in Section III, we found that fine-grained passenger demand models experience uncertainty. Thus, we formulate the passenger demand $\beta_{j\cdot d}$ in our optimization as a variable, instead of a fixed value. In this work, we assume $\beta_{j\cdot d}$ belongs to some uncertainty set where $\underline{\beta}_{j\cdot d} \leq \beta_{j\cdot d} \leq \overline{\beta}_{j\cdot d}$. As a result, we formulate a robust optimization problem as follows.

Definition 3 (**Robust Advertising Efficiency Optimization**).

$$\max_{\lambda_{j\cdot d}^i} \min_{\beta_{j\cdot d}} \frac{\sum_{i=1}^{\gamma}\sum_{d=1}^{m}\sum_{j=1}^{\frac{T}{\tau}} \beta_{j\cdot d} \times \lambda_{j\cdot d}^i}{\sum_{d=1}^{m}\sum_{i=1}^{\gamma} \alpha_d^i}$$

$$s.t.\ \tau \times \sum_{j=1}^{\frac{T}{\tau}} \lambda_{j\cdot d}^i = \alpha_d^i, \forall i \in [1,\gamma], \forall d \in [1,m] \qquad (4)$$

$$\beta_{j\cdot d} \in [\underline{\beta}_{j\cdot d}, \overline{\beta}_{j\cdot d}], \forall j \in [1, \frac{T}{\tau}], \forall d \in [1,m]$$

In this problem, the key challenge is how to obtain an uncertainty set of passenger demand, i.e., $\underline{\beta}_{j\cdot d}$ and $\overline{\beta}_{j\cdot d}$ and then to determine the schedule $\{\lambda_j\}_{j=1}^{\frac{T}{\tau}}$ with an online fashion, so that the best n slots with highest passenger demand can be selected as the advertising slots based on the latest info. As follows, we conduct an empirical study in Section III before we present our demand model and scheduling.

III. Empirical Study

To predict future passenger demand, we conduct an extensive empirical study on the real-world AFC records collected in the metro system in Shenzhen, China.

A. Dataset

In this paper, we utilize two sample datasets from a streaming dataset of smartcard transactions in the metro AFC system of Shenzhen, which is the twin city of Hong Kong and has more than 13 million residents. Each card swiping record includes card ID, device ID, swiping in or out, date, time as well as metro station ID among 118 metro stations. In this paper, two sample datasets, named as Dataset A and Dataset B, are used. The summary of Dataset A and Dataset B as well as the format of each data record are illustrated in Fig. 2. As shown, Dataset A contains 2,807,973 smart cards, and the records range from Dec 12, 2013 to Dec 25, 2013. The average daily number of card swiping in the metro system is more than 3 million. The average daily traffic per station is about 25,000. Dataset B contains the streaming card swiping data of 16,000,000 passengers from July 1, 2011, including smartcards used in both metro systems and bus systems in Shenzhen.

	Dataset A	Dataset B
Collection Period	13/12/12- 13/12/25	11/07/01- Now
Number of Cards	2,807,973	16,000,000
Number of Records	41,270,178	6,212,660,742
Format: Card ID, Station ID, Device ID, Date &Time, Swipe in/out		

Figure 2. The dataset summary and the record format.

We utilize Dataset A as a sample to study all metro stations during a two week period. Dataset B is used for the large-scale evaluation. In the rest of this section, although similar observations are found throughout all the 118 stations, we focus on our empirical study results over the Xixiang station, which is a representative commute station.

B. Certainty of Coarse-grained Demand

We first study the trend of coarse-grained passenger demand in terms of traffic volumes shown in Fig. 3, which plots the trends of a weekday (polyline) and the nearby historical nine weekdays (bars). The temporal granularity of the traffic volume is one hour, which is coarse-grained compared to the typical duration of an advertisement, i.e., one minute.

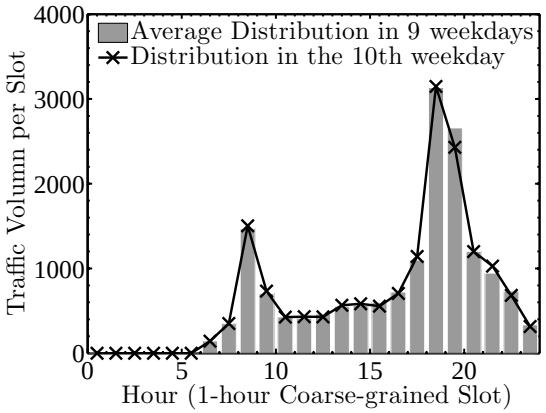


Figure 3. The distribution of the coarse-grained passenger demand in a weekday and the average distribution of its near historical weekdays.

Observation 1. *Predictable* and *Smooth* Coarse-grained Passenger Demand: The trend of the single weekday highly matches the trend of its near historical weekdays. The curve is smooth in general, except for the rush hour. In other words, the trend of the coarse-grained passenger demand in a station is predictable based on the nearby history records.

The main reason for the coarse-grained predictability and smoothness is that the passenger demand is relatively stable during near days, in a specified area around a station. Based on Observation 1, we conclude that the distribution of coarse-grained passenger in a station can be accurately predicted using the average history distribution.

C. Uncertainty of Fine-grained Demand

We now introduce the trend of the fine-grained passenger demand. Fig. 4 plots the distribution of fine-grained passenger demand during the rush hour (from 18:30 to 19:00) of a weekday and its nearby historical nine weekdays.

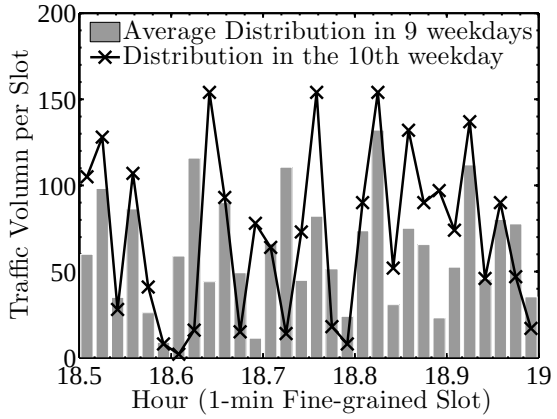


Figure 4. The distribution of the fine-grained traffic volumes in the rush hour (18:30 to 19:00) in a weekday and its near historical weekdays.

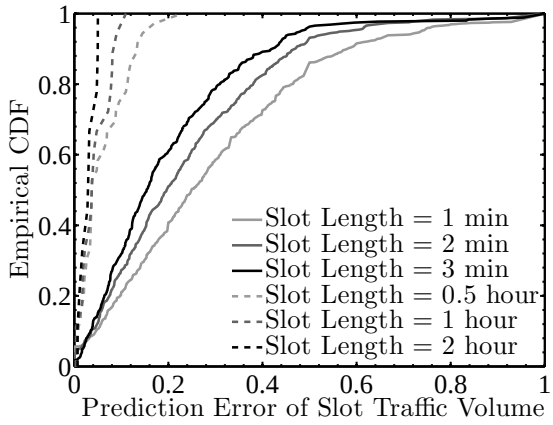


Figure 5. The CDFs of the prediction error of both coarse-grained and fine-grained traffic volumes in a weekday based on the near history.

Observation 2. *Unpredictable* and *Fluctuated* Fine-grained Passenger Demand: The trends of the fine-grained traffic volume in the weekday and its nearby historical weekdays do not match, e.g., the traffic volume in the same slot largely varies in different days. Further, the fine-grained traffic volume, even during rush hours with crowded passengers, is fluctuated with continuous peak and valley values, e.g., some valleys next to a peak even have a traffic volume as low as the average volume in slack hours.

For the unpredictability of fine-grained passenger demand, there are two main reasons. On the one hand, the trains arriving at the same station might not be punctual. On the other hand, although trains arrive punctually, a commuter cannot guarantee to take the same train every day even the commuter needs to arrive at the work place before 9am every day. The arrivals of the passengers at a given metro station are temporally different in different days. As a result, we argue that a higher historical traffic volume cannot guarantee higher advertising efficiency for fine-grained advertising. This motivates us to predict fine-grained demand by inferring real-time individual arrivals, instead of relying on pure history.

For the fluctuation of fine-grained passenger demand, the reason is that the passenger arrivals are highly limited by the schedule of metro trains. Between the arrivals of two successive trains in a metro station, no passengers arrive at the station. Inspired by this observation, we aim to propose fine-grained advertising instead of coarse-grained advertising because the existence of valleys slots during a coarse-grained period (e.g., rush hours) pulls down the advertising efficiency of that period. Only the fine-grained selection of all the high-traffic slots leads to better advertising efficiency.

D. Historical Prediction Performance

We now reveal the different performance of the historical prediction of fine-grained and coarse-grained passenger demand. If the slot traffic volume of t_j, i.e., β_j, is predicted as the average volume of t_j over the near historical nine weekdays, denoted by $\hat{\beta}_j$, the prediction errors when slot length varies are shown in Fig. 5. The prediction error, denoted by δ_j, is computed by $\delta_j = \frac{|\beta_j - \hat{\beta}_j|}{\max\{\beta_j, \hat{\beta}_j\}}$. As expected, the prediction error of coarse-grained traffic volume is much lower than that of fine-grained traffic volume. For example, when the slot length is 1 h, the prediction errors of around 99% slots are lower than 10%. However, more than 40% slots have a prediction error larger than 30% when the slot length is 1 min.

E. Summary

Based on the empirical study, we have following important conclusions: (i) According to Observation 1, the coarse-grained passenger demand is predictable based on historical distributions, which provides us a global view on the coarse-grained passenger demand in a future day. (ii) According to Observation 2, the fine-grained passenger demand is unpredictable based on historical distributions. As a result, we propose an uncertain passenger demand model to predict fine-grained traffic volumes based on individual arrivals prediction. (iii) According to Observation 2, the fine-grained, instead of coarse-grained, advertising scheduling should be applied due to fluctuated fine-grained passenger demand. Thus, the selection of peak fine-grained slots among all the fluctuated slots is enabled to achieve higher advertising efficiency. So we design an online advertising scheduling based on robust receding horizon control. As follows, we first give an overview of our approach in Section IV, and then introduce our demand model and AD scheduling at Sections V and VI, respectively.

IV. Methodology Overview

Inspired by the empirical study, we propose the "Everyone Counts" design to improve the advertising efficiency by robust receding horizon control based on a two-level passenger demand model. As follows, we introduce the overview of the proposed approach. In the rest of the paper, we use **frame**, denoted by $\{f_l\}_{l=1}^{\frac{T}{h}}$, to represent the coarse-grained time slot whose length, denoted by h, is in hour level. We use **slot**, with minute-level length τ, denoted by $\{t_j^l\}_{j=1}^{\frac{h}{\tau}}$, to represent the jth fine-grained slot in frame f_l. Our demand modeling and AD scheduling are based on these two temporal units.

Given the fixed advertising time α_d^i and slot length τ, the proposed approach (i) allocates the total $n = \alpha_d^i/\tau$ AD slots

to coarse-grained frames based on our *coarse-grained demand model* obtained by historical AFC data, and (ii) allocates all AD slots belonging a particular frame based on our *fine-grained demand model* obtained by real-time AFC data plus individual mobility patterns.

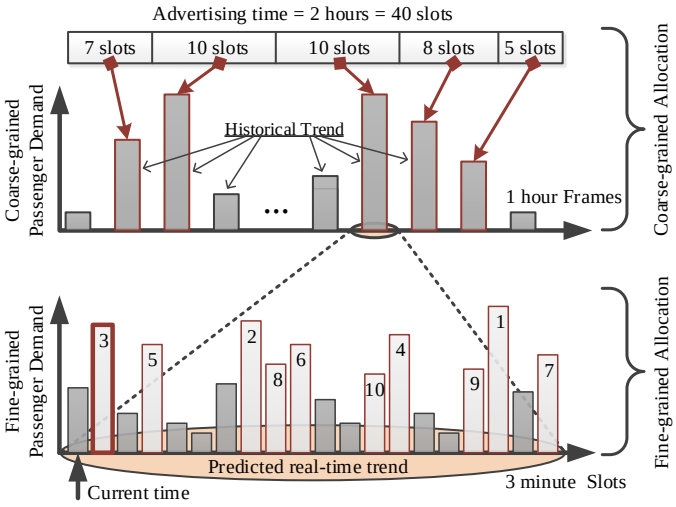


Figure 6. An example illustrating the overview of the proposed approach.

We use an example regarding to a specific metro station ψ_d and a specific advertiser i to illustrate the main process of the proposed approach, since scheduling at different stations is independent among each other. As shown in Fig. 6, the length of the advertising time α_d^i is two hours, and the length of slots τ is 3 minutes. Then the total number of advertising slots (AD slots) n is 40. The length of frames h is 1 h. Thus, there are 24 frames in the day and 20 slots in each frame. Our scheduling has two key steps. (i) According to the historical distribution of coarse-grained passenger demand in frames, n AD slots are allocated to $\{f_l\}_{l=1}^{24}$. Note the allocation based on the global view of high-traffic slots distribution in frames enables frame-scale (instead of day-scale) passenger demand prediction, which leads to higher demand prediction accuracy. (ii) Taking one of the frames with 10 AD slots for example, fine-grained passenger demand in this frame is predicted online and updated in each time slot with a receding horizon of newly occurred AFC records. Suppose the best 10 slots with high traffic volumes are those illustrated in the figure, according to the prediction at the current time (i.e., the first slot in the frame). These best 10 slots may be replaced by other slots in further predictions with a receding horizon. Based on this slot-level demand model, an online AD schedule is made for all the slots in the current frame, and the schedule is implemented only for the nearest future slot, i.e., the next slot of the current one. In this example, the next slot ranks 3rd among all the slots in terms of traffic volume, which will be selected as an AD slot. The online prediction and scheduling continue until all the slots in the frame are scheduled.

We next present the detailed design of our demand model and AD scheduling in the following two sections, respectively.

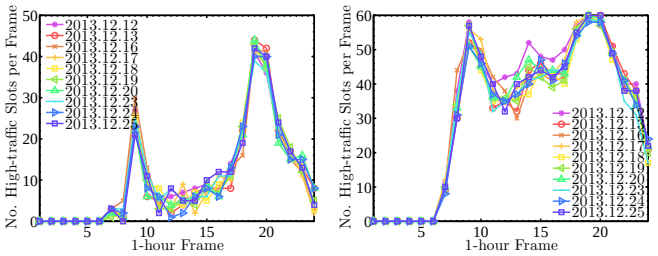
V. Uncertain Demand Model

Our uncertain demand model has two parts, i.e., coarse-grained passenger demand modeling at frame levels, and fine-grained passenger demand modeling at slot levels.

A. Coarse-grained Passenger Demand Modeling

Given historical AFC data, our coarse-grained passenger demand model is a statistical model, i.e., at frame-levels using the historical average passenger demand as the future passenger demand. Based on setting, our coarse-grained frame-level passenger demand model has the following property.

Property 1: Stable Distribution of High Demand Slots in Frames. When all the time slots in a day are sorted according to their passenger demand, the distribution of those best n slots with highest demand in each frame is stable.



(a) Distribution of the best 240 slots with highest traffic volumes

(b) Distribution of the best 720 slots with highest traffic volumes

Figure 7. The distribution of high-traffic slots ($\tau = 1$ min) in each 1 h frame when the advertising time α is 4 h or 12 h. This result is based on Dataset A.

Fig. 7 shows the distributions of the best n slots in frames of short term historical days when h and τ are 1 hour and 1 minute, respectively. Fig. 7(a) plots the distribution of the best 240 slots, i.e., 4 hour advertising time, and Fig. 7(b) plots the distribution of the best 720 slots, i.e., 12 hour advertising time. From the figures, we found that no matter n is large or small, the distribution of the best n slots in frames is stable.

B. Fine-grained Passenger Demand Modeling

Given both historical and real-time AFC data, our fine-grained passenger demand at slot levels is a bayesian model, i.e., based on passenger real-time entering AFC records and mobility patterns, we infer passenger exiting stations and time. Since the passenger demand in a future slot is the number of passenger arriving at the target station from other metro stations, their AFC records entering the metro system have already been recorded. We use such logged AFC records, including both the origin station and the entering time, as a condition to predict passenger arrivals, thus to obtain passenger demand at slot levels. In detail, there are three steps, namely, destination prediction, travel duration prediction, and traffic volume aggregation.

1) Destination Prediction: For a passenger k entering the metro through station θ_k at the time μ_k^θ, we predict the destination d_k based on the recorded AFC entering information according to the predictable passenger destinations.

Property 2: Low-entropy Destination. The entropy of destination, $H(d_k)$, for a metro passenger k is low and the conditional entropy $H(d_k|\theta_k, \mu_k^\theta)$ is even much lower when the origin θ_k and starting time μ_k^θ are given.

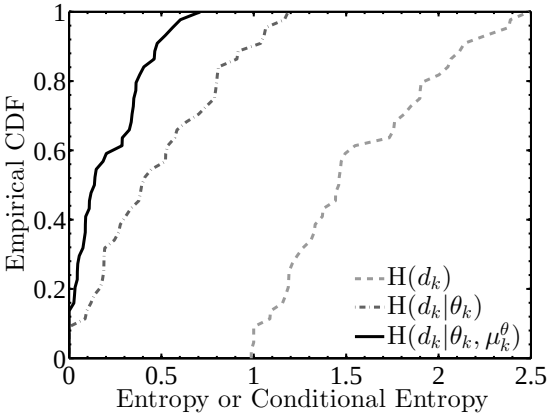


Figure 8. The CDFs of the entropy of destinations and the conditional entropy of the destinations given origins and the hour of trip starting time. The result is based on Dataset B.

Fig. 8 plots the CDFs of $H(d_k)$, $H(d_k|\theta_k)$, and $H(d_k|\theta_k, \mu_k^\theta)$ for all the metro records of passenger k during 3 months, which are computed as follows,

$$\begin{aligned} H(d_k) &= -\sum_{d_k \in \Psi} p(d_k) \log p(d_k), \\ H(d_k|\theta_k) &= \sum_{d_k, \theta_k \in \Psi} p(\theta_k, d_k) \log \tfrac{p(\theta_k)}{p(\theta_k, d_k)}, \\ H(d_k|\theta_k, \mu_k^\theta) &= \sum_{d_k, \theta_k \in \Psi, \mu_k^\theta \in \chi} p(\theta_k, d_k, \mu_k^\theta) \log \tfrac{p(\theta_k, \mu_k^\theta)}{p(\theta_k, d_k, \mu_k^\theta)}, \end{aligned} \tag{5}$$

where Ψ, the set of all the metro stations, is the support of θ_k and d_k, which can be considered as random variables, and χ = {[00:00:00, 01:00:00), [01:00:00, 02:00:00), $\cdots$, [23:00:00, 24:00:00)}, the set of hours in a day, is the support of the random variable μ_k^θ. We find from the figure that $H(d_k|\theta_k, \mu_k^\theta)$ is lower than 0.7, which means there are only $2^{0.7}$ possible destinations, compared to totally 118 stations in Shenzhen metro, for each passenger when θ_k and the hour of μ_k^θ are given. By exploring the travel history of each passenger, the destination of a future metro trip can be exactly predicted given the recorded AFC information of the origin and the starting time of the trip, i.e., finding the destinations mostly associated with the origin and the starting time in the historical data.

2) Travel Duration Prediction: We then predict the arrival time ν_k^d of passenger k at the predicted destination d_k based on the recorded θ_k and μ_k^θ, according to the following property of passenger mobility.

Property 3: Stable Travel Duration. The travel durations between a given pair of origin and destination, denoted by $\omega(\theta, d)$, is stable. Moreover, if the hour of the starting time μ^θ is given, the standard deviation of the travel durations for a pair of origin and destination, denoted by $\sigma(\omega(\theta, d))$, will be further reduced.

Fig. 9 plots the CDFs of $\sigma(\omega(\theta, d))$, and $\sigma(\omega(\theta, d|\mu^\theta)), \theta, d \in \Psi$, where $\omega(\theta, d|\mu^\theta)$ represents

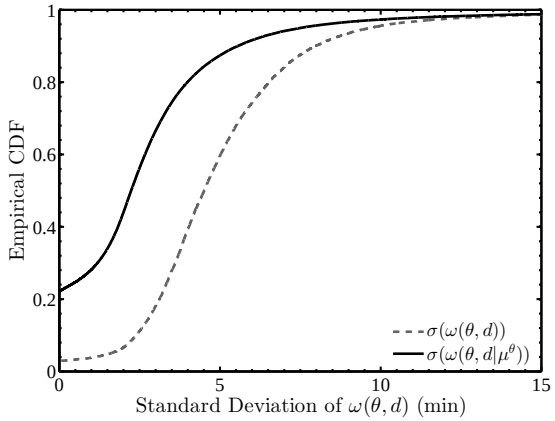


Figure 9. The CDFs of the standard deviation of travel durations given a pair of origin and destination and the hour of trip starting time. This result is based on Dataset B.

the travel durations for a given pair of stations and starting during a given hour in χ. We find from the figure that 70% of $\sigma(\omega(\theta, d|\mu^\theta))$ are lower than 3 min.

In this work, we use 90th percentile of the standard deviation to obtain an arriving time interval, instead of a fixed value, for an uncertain demand model.

3) Traffic Volume Aggregation: Based on the arriving time intervals of all passengers who are already in the metro network, we obtain a lower bound, i.e., $\underline{\beta}_{j \cdot d}$, and a upper bound, i.e., $\overline{\beta}_{j \cdot d}$ for passenger demand at station ψ_d during slot t_j based on all the AFC records occurring before the current slot t_{j-1}. These lower and upper bounds are used in our scheduling with robust receding horizon control as follows.

VI. Scheduling by Receding Horizon Control

Based on our two-level demand model, our scheduling also has two parts, i.e., offline advertising time allocation at frame level, and online advertising time allocation at slot level.

A. Offline Advertising Time Allocation at Frame Level

Given n AD slots and the historical coarse-grained traffic distribution of $\frac{T}{h}$ frames, we first allocate n to the frames. The number of AD slots for frame $f_l, 1 \le l \le \frac{T}{h}$, is denoted by η_l. The allocation is inspired by the Property 1 of our coarse-grained demand model, i.e., stable distribution of high demand slots in frames. So we determine the allocation $\{\eta_l\}_{l=1}^{\frac{T}{h}}$ according to the stable distribution of high-traffic slots in frames. Given the fixed n and the historical percentage, denoted by π_l, of the high-traffic slots in frame f_l, η_l is computed as the product of n and the stable historical percentage π_l, i.e., $\eta_l = n \times \pi_l$, where $\sum_{l=1}^{\frac{T}{h}} \eta_l = n$, and $\sum_{l=1}^{\frac{T}{h}} \pi_l = 1$.

B. Online Advertising Time Allocation at Slot Level

With the allocation of the number of the advertising slots in each frame and the predicted real-time traffic volumes in future slots in each frame, the schedule $\{\lambda_j^l\}$ are made with receding horizons of real-time AFC records, where $1 \le l \le \frac{T}{h}, 1 \le j \le \frac{h}{\tau}$. The pseudo code in Alg. 1 explains the main process of the scheduling.

Algorithm 1: Online AD Slot Allocation

Input $\{\eta_l\}$: the allocation of n advertising slots for each frame f_l; Real-time card swiping records
Output $\{\lambda_j^l\}$: the schedule of all the time slots in all the frames

```
l = 1, j = 0
for l = 1, l <= T/h (i.e., all the frames) do
    for j = 0, j < h/τ do
        Predict the lower and upper bounds for
        {β^l_{j+1}, β^l_{j+2}, ··· , β^l_{h/τ}} based on the updated AFC records
        occurring before t^l_j.
        Solve robust advertising efficiency optimization problem
        proposed in eq.(4).
        When the current time moves into t^l_{j+1}, display
        advertisements if λ^l_{j+1} is 1.
        Continue
    end for
end for
```

During a frame f_l, before each slot t_j^l, the advertising schedule λ_j^l, is made by first predicting lower and upper bounds of passenger demand in the future slots in f_l, $\{\beta_{j+1}^l, \cdots, \beta_{\frac{h}{\tau}}^l\}$ with updated AFC records. Then, we solve the robust advertising efficiency optimization problem proposed in eq.(4) by a numerical method to obtain the schedule for the rest of slots. Then, we choose to display the AD or not based on the obtained schedule.

VII. Performance Evaluation

Based on one month data from Dataset B in Fig. 2, we evaluate the performance of the advertising time allocation, the fine-grained traffic prediction, and the resulted advertising efficiency of our approach against different settings of the slot length (default: 3 min), frame length (default: 1 hour), and the length of advertising time (default: 5 hours).

We compare the performance of the proposed approach with the performance of the optimal scheduling and two existing approaches as follows.

- **The optimal approach (Optimal):** This approach makes the optimal schedule for each time slot to achieve the optimal advertising efficiency under the given advertising time and slot length, based on the ground truth of fine-grained traffic volumes.
- **Coarse-grained scheduling approach (Coarse-grained):** This approach selects the best coarse-grained frames with highest historical traffic volumes in near historical days to display advertisements. The frames are sorted and selected to display advertisements in decreasing order of their historical frame traffic volume.
- **Historical traffic volume based approach (Historical):** This approach applies fine-grained advertising based on historical traffic volume instead of individual mobility pattern based traffic volume prediction. Fine-grained slots are sorted and selected as advertising slots in decreasing order of the historical slot traffic volume in near history.

A. Performance of Advertising Time Allocation

We first explore the allocation accuracy of the advertising slots of EveryoneCounts under different settings of the slot length τ and the frame length h, given the default value of α.

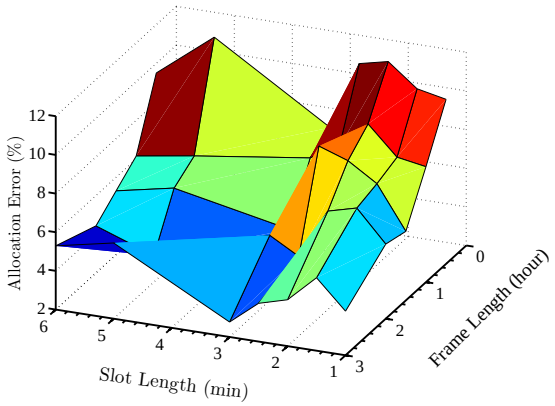


Figure 10. The allocation error of the given advertising time (five hours) into coarse-grained frames when both the length of slots and the length of frames vary.

Fig. 10 plots the allocation error δ_η when τ varies from 1 min to 6 min and h varies from 0.5 hour to 3 hour. The allocation error δ_η is computed as the average prediction error of the number of high-traffic slots in each frame,

$$\delta_\eta = \frac{h}{T} \sum_{l=1}^{\frac{T}{h}} \frac{|\eta_l - \hat{\eta}_l|}{\max\{\eta_l, \hat{\eta}_l\}}, \tag{6}$$

where $\{\hat{\eta}_l\}$ are the true distribution of the high-traffic slots in the frames of the target day.

We can find from the figure that the allocation error of most settings are lower than 10%, especially when the slot length is 3 min and the frame length is longer, e.g., 3 h. The reason for the largely decreasing allocation error as the frame length increases is that the more coarse-grained traffic volume is more predictable. This phenomenon corresponds to the Observation 1 that coarse-grained traffic volumes are more stable than fine-grained traffic volumes.

We can also find that when the slot length is 3 min, the allocation error is the lowest among those with the same frame length. The key reason is as follows. The train interval in the metro system in Shenzhen city is 3 min in rush hours and 6 min in nonrush hours. Then, the 3 min time slots naturally grouping the passengers getting off from the same train. As a result, the traffic volumes in every 3 min slot are more regular than in other length of slots. Inspired by this conclusion, we suggest the setting of slot length τ according to the train interval of the target metro station.

B. Performance of Traffic Prediction

We then evaluate the performance of the online fine-grained traffic volume prediction of our approach by comparing the prediction error with that of Historical under different settings of the slot length, which are plotted in Fig. 11.

We can find from the figure that the prediction error of both approaches decrease as the slot length increases and the prediction error of EveryoneCounts is on average 61.49% lower than that of Historical. When the slot length is as short as 1 min, the prediction performance of both approaches is poor while the reasons of their poor performance are different.

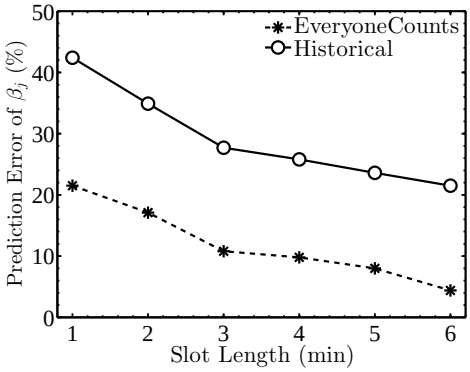


Figure 11. The prediction error of fine-grained traffic volumes of the proposed approach, compared to the history based approach.

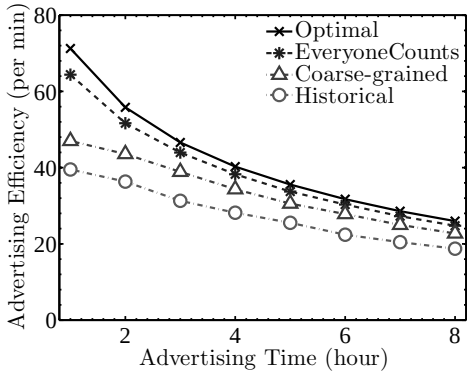


Figure 13. The advertising efficiency of all the approaches when the advertising time varies from 1 h to 8 h.

The reason for the high prediction error of EveryoneCounts with 1 min slot length is the deviation of the individual travel duration from the average historical duration of each passenger (as shown in Fig. 9) is much larger than the 1 min slot length. While the reason for the extremely high prediction error of Historical is the uncertainty of the fine-grained traffic volume, our novel observations made in Section III-C.

C. Performance of Advertising Efficiency

The performance of the advertising efficiency of all the approaches are evaluated against the impact of advertising time α, slot length τ and frame length h.

1) Impact of the Advertising Time: Fig. 12 and Fig. 13 plot the total traffic volume accumulated during all the advertising time (the objective of the equivalent problem presented in Equation (4)) and the advertising efficiency when the given advertising time varies from 1 h to 8 h.

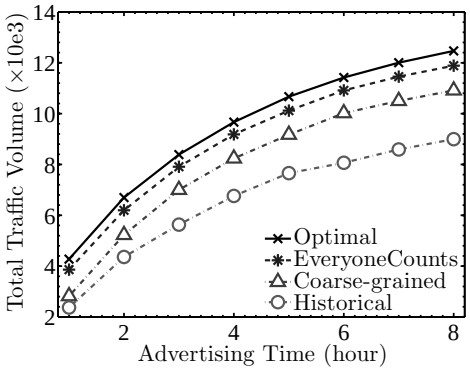


Figure 12. The total traffic volumes of all the approaches when the advertising time varies from 1 h to 8 h.

From the two figures, the proposed approach performs the best among all the non-optimal approaches. As the advertising time increases, as expected, the total traffic volume increases, while the advertising efficiency decreases for all the approaches. Since the passenger traffic is not evenly distributed in the target day and those time slots with higher traffic volumes will be preferentially selected, the longer the advertising time, the lower the advertising efficiency. When the advertising time is 1 h, the total traffic volume (advertising efficiency) of EveryoneCounts is 37% and 62.9% higher than that of Coarse-grained and that of Historical, respectively. The reason for the lower advertising efficiency of Coarse-grained than that of EveryoneCounts is the fluctuated fine-grained traffic volume even during rush hours, which has been introduced as Observation 2. Historical has the worst performance of the advertising efficiency because of the high prediction error of the fine-grained traffic volume, which is resulted by the temporal irregularity of fine-grained traffic volumes.

2) Impact of the Slot Length: Given the fixed length of the advertising time, the length of the slots has different impacts on the advertising efficiency of different approaches. Fig. 14 plots the advertising efficiency of all the approaches when the slot length varies from 1 min to 6 min.

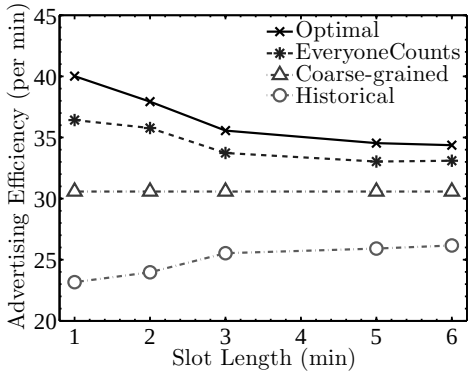


Figure 14. The advertising efficiency of all the approaches when the slot length varies from 1 min to 6 min.

Optimal and our approach have decreasing advertising efficiency while Historical has increasing advertising efficiency as the slot length increases. For Coarse-grained, since it applies the coarse-grained scheduling, the slot length has no impact on its performance. The decreasing advertising efficiency of Optimal and EveryoneCounts is resulted by the more coarse-grained advertising with longer slot length. Although EveryoneCounts has a decreasing advertising efficiency, the distance between its efficiency and that of Optimal decreases. For example, when the slot length is 1 min and 6 min, the efficiency of EveryoneCounts is 91% and 96.3% of that of Optimal, respectively. As expected, the lowest advertising efficiency of Historical when the slot length is shortest is resulted from the temporal uncertainty of fine-grained traffic volume.

3) Impact of the Frame Length: Fig. 15 presents the advertising efficiency of approaches when the frame length varies

from 0.5 h to 3 h. Optimal and Historical are not affected by the frame length because they both apply only fine-grained advertising. Our approach and Coarse-grained have decreasing performance with larger frame length because they both take advantage of the coarse-grained traffic certainty. However, longer frame length leads to higher traffic volume prediction error of EveryoneCounts because that more and farther future traffic volumes need to be predicted in longer frames.

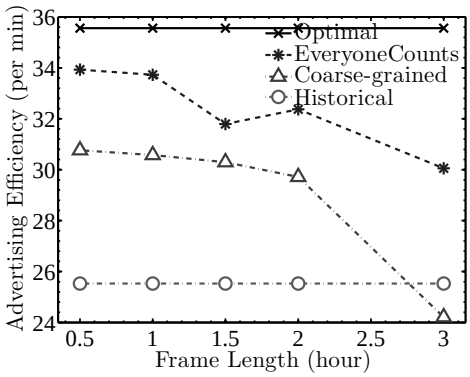


Figure 15. The advertising efficiency of all the approaches when the frame length varies from 0.5 hour to 3 hour.

4) Global Performance: We show the advertising efficiency achieved by all the approaches in all the metro stations under the default setting in Fig. 16. From the figure, we can find that EveryoneCounts has the closet performance with Optimal. The average advertising efficiency of EveryoneCounts is 26.84, which is 89.7% of the efficiency of Optimal, and 18.6% and 58.5% higher than the efficiency of Coarse-grained and Historical, respectively.

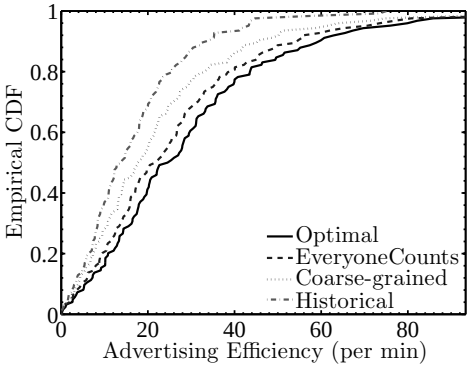


Figure 16. The empirical CDFs of the advertising efficiency in all the 118 metro stations of all the approaches under the default setting.

VIII. Discussions

In this section, we discuss some practical issues regarding to the real-world deployment of the proposed approach.

Advertisement Fairness: The proposed approach is to assist a metro system optimize its advertising efficiency given the total advertising time rented by all advertisers. Since an advertiser is charged according to the length of its advertising time, it is also important to maintain the fairness among all advertisers. Our approach can be easily adapted to take the fairness into account. When a time slot is selected as an advertising slot, each advertisement will be displayed with a probability proportional to the its length of rented time. As a result, when the rented time of an advertisement is longer, its display probability would be higher.

Passengers Contributing to Revenues: In this paper, we envision that all passengers passing by a digital screen would see advertisements. Admittedly, in reality, only a portion of passengers would see advertisements, and an even smaller portion of passengers would actually contribute to advertisers' revenues. However, these real-world factors are extremely difficult to quantify, and are out of the technical scope of this paper. Thus, we focus on the total passing by passenger traffic volume in this work.

Passengers with Temporary Cards: In a real-world metro system, there are passengers traveling with temporary cards, e.g., foreign visitors. Thus, the mobility patterns of these passengers cannot be inferred from the historical travel records. In this work, for these passengers, the destinations and the travel durations are predicted based on statistical general regularities of the passenger mobility within the metro system.

Metro Station Exits Without AFC Machines: Some metro stations might not have AFC machines at the exits, so the destinations of passengers are unavailable. Our approach is still applicable for such stations, because the destination and travel duration of a passenger trip can be estimated by exploring the entering record of the passenger's next trip [4].

Demand Modeling using Other Infrastructures: Other infrastructures, e.g., wifi routers or infrared sensors, can also be used to model passenger demand. But they typically introduce additional costs. In contrast, our approach based on AFC data did not introduce any additional costs, since AFC data are collected automatically for billing purposes.

IX. Related Work

Our approach predicts the passenger traffic volume by utilizing the real-world card-swiping information recorded by the metro automatic fare collection (AFC) system. The most related topics with this work include advertising efficiency improvement, traffic volume prediction, and traffic arrival time prediction. Though smart card data have been used before [3], our work is the first one on the advertising efficiency improvement based on an uncertain demand model. To the best of our knowledge, there is little, if any, research on the topic, so we summary our related work within the area of traffic arrival time prediction and passenger demand estimation.

A. Arrival Time Prediction

In our design, we predict the passenger arrivals in metro networks, which is determined by the arrivals of metro trains. Similarly, several existing studies propose wise designs to predict the travel time of other kinds of transportation, e.g., bus. We classify such work into three categories according to the different information they use, i.e., (i) road sensors, (ii) the taxi GPS information [5], [6], and (iii) the cell phone data [7].

Several pieces of work have been proposed to use road sensors, e.g., the loop detectors, to predict travel time. Based on the data collected by the road sensors, such approaches

predict the travel time by estimating speeds of vehicles. Given the estimated vehicle speed as well as the fixed length of the road segments, these studies further predict the travel time. Recently, more researchers cast their attentions on the large scale taxi GPS data. Balan et al. [5] propose a real-time trip information system that provides passengers with the expected fare and travel time. The authors in [6] propose a citywide model for estimating the travel time of any path in real time, based on the map information and the GPS trajectories of vehicles received in current time slots as well as the history records. Zhou et al. [7] propose a novel idea to predict the bus arrival time using the cell phone data. They present a bus arrival time prediction system based on the participatory sensing data provided by cell phones of bus passengers. The work VTrack [8] also uses sensing data from phones, e.g., the WiFi-based positioning samples, to predict traffic delay.

B. Passenger Demand Estimation

The existing work on the real-time passenger demand, i.e., traffic volume estimation, mainly focuses on the volume estimation of vehicles, which can be classified into two categories, i.e., parametric and non-parametric methods. Parametric methods typically used models include local regression model [9], and Markov chain model [10], etc. Non-parametric methods include non-parametric regression [11], Bayesian networks [12] and neural networks [13].

Based on the source of the data used for prediction, the related work can also be divided into two categories, i.e., (i) the approach using road sensors [8], [14], [15], and (ii) the approach using taxi GPS data [16]–[21]. Singliar et al. [14] develop a probabilistic estimation model for highway networks based on the information collected from a set of traffic sensors placed around the city. The authors in [16] use both historical patterns and real-time traffic information from the GPS data of taxicabs to estimate traffic conditions. Aslam et al. [17] provide model and inference procedures which can be used to analyze traffic patterns from historical data, and to estimate current traffic status from data collected in real-time. Yuan et al. [21] propose a taxi passenger demand model for taxi drivers to quickly pick up passengers to maximum their revenue.

C. Summary

Our work solves the advertising optimization problem based on our uncertain passenger demand model with card-swiping records collected by AFC machines in a metro system. Compared with other kinds of transportation, e.g., cars or buses, the metro system has quite different properties in both metro train traveling and passenger mobility patterns. The existing approaches for vehicle traffic prediction cannot be directly used to predict metro passenger demand.

X. Conclusion and Future Work

In this paper, we propose a novel online approach for metro digital advertising to improve advertising efficiency. Our extensive empirical study and technical efforts provide a few valuable insights on metro transit networks, which are hoped to be useful for fellow researchers on similar topics. Specifically, we find that (i) coarse-grained passenger demand is regular while find-grained passenger demand is more dynamic; (ii) passenger mobility patterns are fairly stable, and given entering station and entering time, the exiting station can be predicted with a high accuracy; (iii) travel periods between same stations are also stable in different time; (iv) it has a high accuracy to predict passenger demand by using time and stations passengers entering the metro network as conditions.

XI. Acknowledgements

This work was supported in part by the Shanghai Recruitment Program of Global Experts, NSF CNS-1446640, NSFC 61573245, and NSFC 61174127.

References

[1] *New Screens Bring Added Revenue to the MTA*, http://new.mta.info/news/2014/03/13/new-screens-bring-added-revenue-mta.

[2] *Subway and Bus Ridership*, http://web.mta.info/nyct/facts/ridership/index.htm.

[3] N. Lathia and L. Capra, "How smart is your smartcard?: Measuring travel behaviours, perceptions, and incentives," ser. UbiComp '11.

[4] M. Trepanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *Journal of Intelligent Transportation Systems*, 2007.

[5] R. K. Balan, K. X. Nguyen, and L. Jiang, "Real-time trip information service for a large taxi fleet," in *ACM MobiSys*, 2011.

[6] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *ACM KDD*, 2014.

[7] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: Predicting bus arrival time with mobile phone based participatory sensing," in *ACM MobiSys*, 2012.

[8] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *ACM SenSys*, 2009.

[9] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of local linear regression model for short-term traffic forecasting," *Transportation Research Record: Journal of the Transportation Research Board Publisher*, 2007.

[10] G. Yu, J. Hu, C. Zhang, L. Zhuang, and J. Song, "Short-term traffic flow forecasting based on markov chain model," in *IEEE Intelligent Vehicles Symposium*, 2003.

[11] T. Zhang, L. Hu, Z. Liu, and Y. Zhang, "Nonparametric regression for the short-term traffic flow forecasting," in *Mechanic Automation and Control Engineering*, 2010.

[12] E. Castillo, J. M. Menndez, and S. Snchez-Cambronero, "Predicting traffic flow using bayesian networks," *Transportation Research Part B: Methodological*, 2008.

[13] F. Jin and S. Sun, "Neural network multitask learning for traffic flow forecasting," in *IEEE International Joint Conference on Neural Networks*, 2008.

[14] T. Singliar and M. Hauskrecht, "Modeling highway traffic volumes," in *ECML*, 2007.

[15] M. Lippi, M. Bertini, and P. Frasconi, "Collective traffic forecasting," in *Machine Learning and Knowledge Discovery in Databases*, 2010.

[16] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *ACM KDD*, 2011.

[17] J. Aslam, S. Lim, X. Pan, and D. Rus, "City-scale traffic estimation from a roving sensor network," in *ACM SenSys*, 2012.

[18] P. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi gps traces," in *Pervasive Computing*, 2012.

[19] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in *KDD '10*.

[20] Y. Huang and J. W. Powell, "Detecting regions of disequilibrium in taxi services under uncertainty," in *SIGSPATIAL '12*.

[21] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in *UbiComp '11*.