IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 13, 2018, accepted June 7, 2018, date of publication June 12, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2846550

A Unified Framework for Predicting KPIs of On-Demand Transport Services

JIHONG GUAN[1], WEILI WANG[1], WENGEN LI[2], AND SHUIGENG ZHOU[3], (Member, IEEE)
[1]Department of Computer Science and Technology, Tongji University, Shanghai 200092, China
[2]Department of Computing, The Hong Kong Polytechnic University, Hong Kong
[3]School of Computer Science, Fudan University, Shanghai 200433, China

Corresponding author: Wengen Li (cswgli@comp.polyu.edu.hk)

This work was supported in part by the Natural Science Foundation of China under Grant 61772367 and in part by the Program of Science and Technology Innovation Action of Science and Technology Commission of Shanghai Municipality (STCSM) under Grant 17511105204.

ABSTRACT Having a better understanding of the key performance indicators (KPIs, e.g., demand and unmet demand) in the next time slot (e.g., next hour) is important for on-demand transport services, such as Uber and DiDi, to improve the service quality. In addition to the spatio-temporal dynamics, KPIs of on-demand transport services are also affected by many exogenous factors from different domains, e.g., the traffic condition from transportation domain and the weather condition from meteorology domain. Therefore, this paper proposes a unified framework to fuse the data collected from different domains to predict multiple KPIs for on-demand transport services. As demonstrated by the experiments, the proposed framework can capture both long-term regularity and short-term dynamics, thus achieving a better performance than the existing solutions in predicting KPIs.

INDEX TERMS On-demand transport service, key performance indicator, cross-domain data fusion, feature selection.

I. INTRODUCTION

In recent years, on-demand transport services such as Uber,[1] DiDi[2] and Lyft[3] have received tremendous popularity due to their remarkable convenience and high efficiency. In on-demand transport services, as illustrated in Figure 1, passengers issue orders while drivers provide services to serve the orders. Different from traditional taxi services, A service platform is introduced to bridge passengers and drivers in on-demand transport services. Concretely, passengers and drivers are connected via an application provided by the service platform. Orders issued by passengers are dispatched to drivers regarding the locations of both orders and drivers, the requirements of orders, and the preferences of drivers. In general, drivers have autonomy to accept or reject the dispatched orders.

For on-demand transport service platforms, it is of high importance to know the Key Performance Indicators (KPIs, e.g., demand, unmet demand, unmet rate and cancel rate of orders) in advance to evaluate the quality of their services in the near future. On the one hand, the dynamic pricing mechanisms in on-demand transport services are highly dependent on these KPIs. For example, the price should increase if the unmet rate is high so that the passengers willing to pay more get a high priority to be served. On the other hand, if the service platforms know the emerging bad KPIs (e.g., a high unmet rate) in advance, they can take appropriate actions to improve the corresponding KPIs to ensure high service quality.

However, it is difficult to achieve accurate prediction for multiple KPIs in on-demand transport services due to the following challenges.

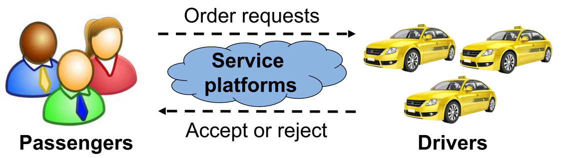


FIGURE 1. The general paradigm of on-demand transport services, where passengers issue orders (i.e., demand) while drivers provide services (i.e., supply) to serve the orders.

[1]https://www.uber.com/
[2]http://www.didichuxing.com/en/
[3]https://www.lyft.com/

- **Cross-domain issue**: The KPIs in on-demand transport services are generally affected by many factors from different domains, such as the traffic condition from transportation domain and the weather condition from meteorology domain. Therefore, it is necessary to propose effective data fusion methods to combine information from different domains to obtain comprehensive knowledge for predicting KPIs. However, the granularity, format and quality of the data from involved domains could be quite different, which makes it challenging to achieve effective cross-domain data fusion.
- **Heterogeneity**: Different KPIs usually have different characteristics. As illustrated by Figure 2, the temporal distributions of demand (i.e., total orders) and unmet rate (i.e., the percentage of orders that are not accepted by any drivers) are significantly different. The demand has high regularity while the unmet rate does not. Therefore, it is difficult to build a unified framework to cover multiple KPIs of different characteristics. Meanwhile, different KPIs are also likely to be correlated. For example, demand is an important factor that affects unmet rate. However, it is generally non-trivial to utilize such correlations to improve the accuracy of prediction due to the heterogeneity of KPIs.

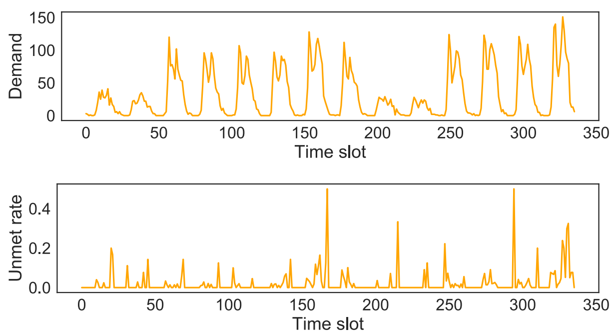


FIGURE 2. The hourly demand and unmet rate of a region in Hong Kong for an on-demand transport service platform that has thousands of regular drivers, where the time period spans 14 days and has 24x14=336 time slots (i.e., hours) in total.

- **High dynamics**: The KPIs of on-demand transport services are dynamically varying over time and space. It is thus difficult to build a prediction model to capture such dual dynamics, i.e., spatio-temporal dynamics.

To address the challenges above, we propose a unified framework which can predict multiple spatially and temporally varying KPIs for on-demand transport services by fusing data from different domains. In practice, data fusion has been widely applied to various spatio-temporal contexts, e.g., decision fusion in wireless sensor network (WSN) [1], [2], and traffic modeling and prediction in mobility-on-demand systems [3]. The major reason for requiring data fusion is that many decisions are related to the information from multiple sources/domains and thus require a data fusing operation to get more comprehensive knowledge. Concretely, in this work, the proposed framework first extracts distinguishable features (e.g., historical KPI features, static urban features, and dynamic urban features) from data in multiple domains. Then, all the extracted features are coupled to train time-dependent KPI prediction models which are further employed to predict KPIs in the future time slots.

In summary, we make the following contributions in this work.

- We formulate the problem of predicting multiple KPIs for on-demand transport services and identify the major challenges to solve this problem.
- We propose a unified framework which extracts features from different domains to predict multiple KPIs and covers both long-term regularity and short-term dynamics in KPIs.
- We conduct extensive experiments to evaluate the performance of the proposed framework by comparing it with the existing methods.

The remainder of this work is organized as follows. Section II reviews the related work on KPI prediction in transport services. Section III presents the definitions of multiple KPIs, formulates the prediction problem, and briefly introduces the datasets involved. Section IV overviews the proposed framework and elaborates the details of its components. Section V conducts experiments to evaluate the proposed solution and provides a case study. We conclude this work and highlight the potential directions for future work in Section VI.

Table 1 lists the notations that will be frequently used in this paper.

TABLE 1. Frequently used notations.

notation	meaning
R_i	a region
T_j	a time slot
o_i	an order
t_o	the response time of order o
$O_{i,j}$	the orders issued in region R_i within time slot T_j
$\|*\|$	the cardinality of set $*$
$\|O_{i,j}\|$	the demand in R_i within T_j
$O^u_{i,j}$	the unmet orders (i.e., unmet demand) in R_i within T_j
$\rho^u_{i,j}$	the unmet rate in R_i within T_j
$O^c_{i,j}$	the cancelled orders (i.e., cancelled demand) in R_i within T_j
$\rho^c_{i,j}$	the cancel rate in R_i within T_j
$p_i{=}(l_i, t_i, s_i)$	GPS point with location l_i, timestamp t_i and status s_i
Tr	a GPS trajectory $Tr = (p_1, p_2, \cdots, p_n)$
I_d	the idle ratio of driver d

II. RELATED WORK

To the best of our knowledge, the research on predicting KPIs for on-demand transport services is quite limited. Due to the lack of readily available data about KPIs, existing studies mainly focus on the *demand prediction* and the *estimation of unmet demand* for traditional taxi services in which drivers cruise along the roads to search for passengers.

A. DEMAND PREDICTION

Demand prediction aims to predict the demand (i.e., the number of orders or passengers) in the future. Roughly, the existing studies dealing with demand prediction can be divided into three categories according to the techniques they use, i.e., *time series-based approaches*, *clustering-based approaches*, and *advanced model-based approaches*.

1) TIME SERIES-BASED APPROACHES [4]–[9]

This line of research takes advantage of the temporal regularity of demand to conduct prediction. Generally, time series analysis techniques like Auto-Regressive Integrated Moving Average (ARIMA) model are used in these approaches. Specifically, Moreira-Matias *et al.* [4]–[6] applied ARIMA model to predict the demand for taxis at each taxi stand; Li *et al.* [7] used ARIMA model to predict the demand for taxis at taxi hotspots; Jiang *et al.* [8] designed an ARIMA-based prediction method to predict the spatio-temporal variation of demand at taxi hotspots. In addition, some researchers also applied Time-Varying Poisson Process [9] to predict the demand for taxis.

2) CLUSTERING-BASED APPROACHES [10]–[12]

Approaches in this category aim to identify hotspots of demand (i.e., the regions or places with much more demand than other regions or places) based on the historical passenger pick-up events. To this end, Liu *et al.* [10] designed a density-based clustering algorithm to find hotspots of demand. To speed up the efficiency of clustering, they projected all data points to a density image and perform clustering on the derived density image instead of the raw data. Chang *et al.* [11] further considered the contexts of time (e.g., morning, noon and afternoon), weather (e.g., raining or not) and location. First, they select the historical orders with the same contexts of the prediction. The selected orders are then clustered based on their locations to identify taxi hotspots. Finally, the generated clusters (i.e., hotspots) are mapped to the road segments to get the semantic meanings. Zhang *et al.* [12] proposed a demand hotspots prediction framework based on spatio-temporal clustering to provide recommendations for taxi drivers. Concretely, they first extract historical pick-up events from GPS trajectories of drivers and divide the extracted pick-up events into 24 time slots (i.e., 24 hours). An adaptive DBSCAN clustering is then conducted on the pick-up events within each time slot to identify the hotspots for that time slot. The identified hotspots are then recommended to drivers according to their locations and the time.

3) ADVANCED MODEL-BASED APPROACHES [13]–[18]

In addition to time series-based approaches and clustering-based approaches, there are also some other approaches for predicting demand using advanced models, e.g., probabilistic model-based approach [13], [14], neural networks-based approach [15], and queueing theory-based approach [16]. Moreover, some studies have also been conducted to discover the patterns of demand in transport services. For example, Bischoff *et al.* [17] investigated the recurrent patterns of demand in Berlin; Lee *et al.* [18] studied the patterns of taxi pick-ups in Jeju, South Korea.

B. ESTIMATION OF UNMET DEMAND

In addition to demand prediction, there are also some studies relevant to unmet demand which usually refers to the orders that are not served by drivers. However, these studies focus on analyzing the degree of unmet demand rather than predict unmet demand directly because it is difficult to collect all passengers' information in traditional taxi services. With the observation that a large number of available taxis indicate a small unmet demand, Afian *et al.* [19] estimated the unmet demand in a region by computing the number of available taxis. Similarly, Shao *et al.* [20] estimated the level of unmet demand by considering how fast an available taxi is occupied after entering a region. Moreover, some researchers studied the relationship between demand and supply by using statistical analysis [21], [22].

C. REMARKS

Our work is different from the studies discussed above in three folds. **First**, existing studies usually focus on demand prediction while we also predict some other KPIs beyond demand, e.g., unmet demand, unmet rate, cancel demand, cancel rate, and average response time. In practice, it is comparatively easy to predict demand due to its high regularity (cf. Figure 2). However, some KPIs are not as regular as demand, which makes existing techniques for demand prediction inapplicable. **Second**, existing studies usually focus on the traditional taxi services in which drivers cruise along the roads to search for potential passengers. Differently, we predict KPIs for on-demand transport services in which passengers and taxis operate via online platforms and do not need to see each other. **Third**, existing studies tend to conduct prediction only based on the historical records of the target KPI (e.g., demand) itself while ignoring other factors. However, KPIs are usually contributed by factors from multiple domains and these factors are indispensable for predicting KPIs accurately.

III. PROBLEM STATEMENT

A. KEY PERFORMANCE INDICATORS

In practice, there are many *Key Performance Indicators* (KPIs) for evaluating the status of a service from different views. Specifically, in on-demand transport services, important KPIs include *demand*, *unmet demand*, *unmet rate*, *cancelled demand*, *cancel rate* and *average response time*. These KPIs collectively reflect the relationship between demand and supply, the quality of services, and the behaviors of passengers and drivers. The definitions of these KPIs are elaborated as below.

1) DEMAND

Given a region R_i and time slot T_j (e.g., from 9am to 10am), we denote all the orders issued in R_i within T_j by $O_{i,j}$.

Accordingly, the demand is denoted by $|O_{i,j}|$, where $| * |$ computes the cardinality of set $*$.

2) UNMET DEMAND AND UNMET RATE

Among all the issued orders $O_{i,j}$ (i.e., demand), those orders that are not accepted by any drivers are regarded as unmet orders and denoted by $O^u_{i,j}$. Therefore, the unmet demand is $|O^u_{i,j}|$ and the corresponding unmet rate $\rho^u_{i,j}$ is computed by

$$\rho^u_{i,j} = \frac{|O^u_{i,j}|}{|O_{i,j}|}. \tag{1}$$

3) CANCELLED DEMAND AND CANCEL RATE

Cancelled demand refers to the orders $O^c_{i,j}$ that are cancelled by drivers or passengers due to various reasons. Cancelled demand $|O^c_{i,j}|$ is an important KPI for analyzing the behaviors of drivers and passengers. With cancelled demand $|O^c_{i,j}|$, the corresponding cancel rate $\rho^c_{i,j}$ is computed by

$$\rho^c_{i,j} = \frac{|O^c_{i,j}|}{|O_{i,j} - O^u_{i,j}|}. \tag{2}$$

where $O_{i,j} - O^u_{i,j}$ computes to those orders that have been accepted by drivers.

4) AVERAGE RESPONSE TIME

The response time of order o refers to the time costed for the order to be accepted by drivers. The average response time t_{avg} of all accepted orders is computed by

$$t_{avg} = \frac{\sum_{o \in O_{i,j} - O^u_{i,j}} t_o}{|O_{i,j} - O^u_{i,j}|}. \tag{3}$$

where t_o is the response time of order o.

B. KPI PREDICTION PROBLEM

In this work, we take an on-demand transport service platform in Hong Kong as an example to present our idea for predicting the frequently used KPIs in on-demand transport services. As illustrated by Figure 3, We partition Hong Kong into 140 small regions according to its administration division.

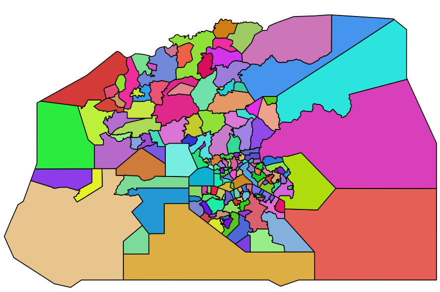


FIGURE 3. The 140 partitions of Hong Kong, where the dense part corresponds to the urban area and the marginal big regions correspond to remote areas.

Specifically, given the historical data from multiple sources/domains, we aim to predict the KPIs like demand, unmet demand, unmet rate, cancelled demand, cancel rate and average response time, for each region in the next time slot (e.g., next hour). Here, the multi-source historical data includes both long-term data and short-term data that will be discussed in Section III-C.

C. DATASETS

Figure 4 illustrates the involved datasets collected from multiple domains, e.g., order data from logistics domain, road network and trajectories from transportation domain, and weather from meteorology domain. These datasets will be used for extracting distinguishable features to predict KPIs. The order data spans consecutive 61 days (two months in 2016) and contains around one million orders. The weather data and trajectory data of drivers for the same period are also collected. Since the region data, road network data and POI data are static, we use the same one for all 61 days.

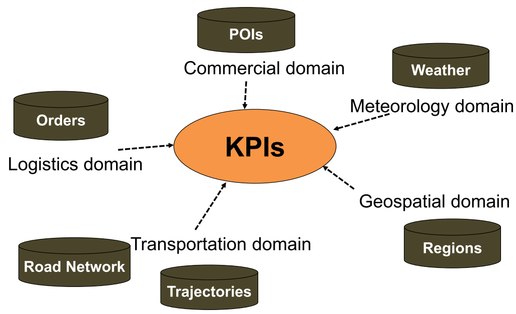


FIGURE 4. Datasets from multiple domains to predict KPIs of on-demand transport services.

The details of these datasets will be discussed in Section IV.

IV. UNIFIED PREDICTION FRAMEWORK

A. OVERVIEW OF FRAMEWORK

Figure 5 presents the proposed unified prediction framework that fuses data from multiple sources to achieve effective KPI prediction. **First**, distinguishable features (e.g., historical KPI features, static urban features and dynamic urban features) are extracted from different datasets. **Second**, since the traffic feature (extracted from driver trajectories) in dynamic urban features is sparse due to the lack of trajectory data, we introduce a traffic inference model to address this sparsity issue based on both static and dynamic urban features. **Finally**, all the extracted features (including KPI features, static urban features, dynamic urban features and inferred traffic feature) and weather feature are fused together to train KPI prediction models. The trained models are then employed to predict KPIs in the next time slot.

All the components of the unified prediction framework will be detailed in the subsequent subsections.

B. EXTRACTING KPI FEATURES

Intuitively, KPIs have certain degrees of temporal regularity, i.e., recurrence/periodicity. For example, the KPIs at the same

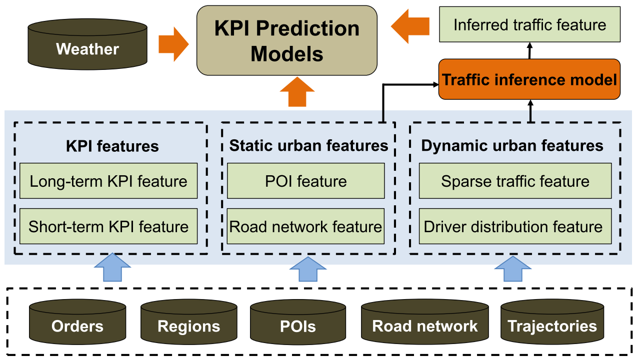


FIGURE 5. **Unified framework for predicting KPIs in on-demand transport services.**

hour of different days could be similar in high probability. Many existing approaches (e.g., Time-Varying Poisson Process [9]) for demand prediction are based on this observation. Meanwhile, the KPIs of adjacent time slots have a high similarity and dependency.

Therefore, for each type of KPI, we compute its historical features from the following two aspects:

- **Long-term KPI feature**: Considering that the same time slots of different days have similar KPIs in high probability, we compute the average value of all the KPIs for each time slot across the whole historical period, i.e., 61 days in this work.
- **Short-term KPI feature**: Considering the similarity and dependency of the KPIs of adjacent time slots, we compute the values of each KPI in previous three time slots. For example, we will consider the KPIs of hours 7, 8, and 9 if we predict the KPI of hour 10. Actually, we evaluated other numbers of time slots and found that three time slots are enough to capture the short-term dynamics and temporal dependency of KPIs.

In this work, the length of each time slot is set to 60 minutes, i.e., one hour for one time slot. However, the proposed framework can be applied to various lengths of time slot without any modifications.

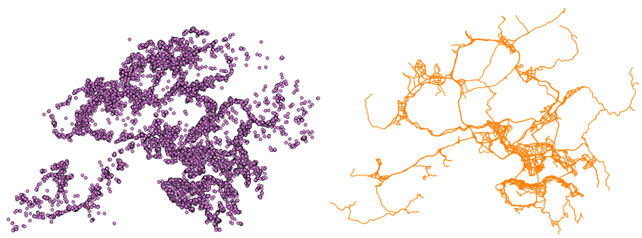
C. EXTRACTING STATIC URBAN FEATURES

KPIs in on-demand transport services are highly dependent on the urban context features like traffic and the distribution of roads. In general, different urban contexts often lead to different demand and supply. For example, the number of orders in the commercial business district is usually larger than that in some remote areas. To capture such urban contexts, we extract both static and dynamic urban features from POI (Point of Interest) data, road network data and trajectory data, where dynamic urban features will be discussed in the next subsection (Section IV-D).

The static urban features, including POI features and road network features, are extracted from two datasets, i.e., POI data and road network data.

1) POI FEATURES

There are many types of POIs, e.g., shops, banks, schools and restaurants, in each region. Figure 6(a) shows the distribution of POIs in Hong Kong. Intuitively, the distribution of POIs can describe the functions of regions to some extent. For example, a commercial area is supposed to have many banks and shopping sites. Figure 7 plots the proportion of the top-15 types of POIs. In this work, we consider the top-10 types of POIs since they already cover around 80% of all the POIs. Therefore, for each region, we compute the total number of POIs and the number of POIs in each type of the ten selected types as the POI features, i.e., 11 features in total.



(a) (b)

FIGURE 6. **The distributions of POIs and road network in Hong Kong. (a) Distribution of POIs. (b) Road network.**

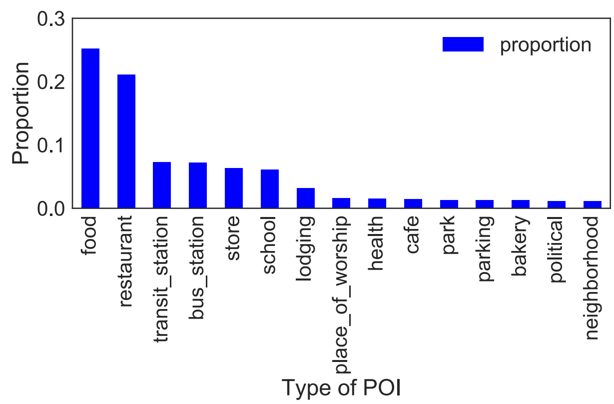


FIGURE 7. **The proportion of the top-15 types of POIs among all POIs.**

2) ROAD NETWORK FEATURES

In addition to POIs, the topology structure of road network is also related to the urban context of one region. Figure 6(b) illustrates the road network of Hong Kong. This road network is extracted from the open source map OpenStreeMap[4] and has a total length of around 1,500 km. There are five major types of roads, i.e., motor way, trunk road, primary road, secondary road and tertiary road, on this road network. For each region, we compute the total length of roads and the length of each type of road as the road network features, i.e., 6 features in total.

[4]https://www.openstreetmap.org/

D. EXTRACTING DYNAMIC URBAN FEATURES

The KPIs of on-demand transport services are also affected by a variety of dynamic urban factors, e.g., the traffic, the distribution of drivers and the weather. These factors tremendously affect the demand and supply, and the behaviors of passengers and drivers. To consider these factors, we extract dynamic urban features from different datasets, where traffic and the distribution of drivers are extracted from drivers' GPS trajectories while weather is extracted from the weather data on Weather Underground.[5]

1) TRAFFIC FEATURE

To obtain the real-time traffic, we compute traffic based on the GPS trajectories of drivers. In the example of on-demand transport service in this work, we have around thousands of regular drivers who report their locations continuously. As illustrated by Figure 8, most sampling time intervals between two location samples within 60 seconds, which provides us abundant information about the traffic.

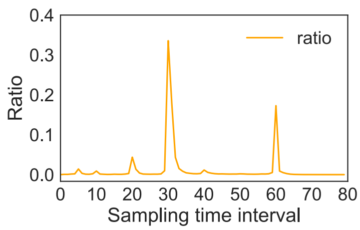


FIGURE 8. The distribution of sampling time intervals among all drivers.

The GPS trajectory Tr of each driver is consisted of a sequence of GPS points, i.e.,

$$Tr = (p_1, p_2, \ldots, p_n) \tag{4}$$

where each GPS point $p_i{=}(l_i, t_i, s_i)$ $(i = 1, \cdots, n)$ contains the spatial location l_i (i.e., longitude and latitude), the timestamp t_i, and the status s_i (i.e., idle or busy) of the corresponding driver. Given two consecutive GPS points p_i and p_{i+1}, the average speed v_i between them is computed by

$$v_i = \frac{dist(l_i, l_{i+1})}{t_{i+1} - t_i} \tag{5}$$

where $dist(l_i, l_{i+1})$ computes the spatial distance between locations l_i and l_{i+1}, and $t_{i+1}-t_i$ computes the time difference between t_{i+1} and t_i. With Eq. (5) and the GPS trajectories of all drivers, we can obtain a set of speed samples $V_{i,j}$ for each region R_i within the specific time period T_j. For region R_i, we compute its average speed $\bar{v}$ and speed deviation σ_v to describe its traffic condition, i.e.,

$$\bar{v} = \frac{\sum_{v \in V_{i,j}} v}{|V_{i,j}|}, \quad \sigma_v = \sqrt{\frac{\sum_{v \in V_{i,j}} (v - \bar{v})^2}{|V_{i,j}|}}$$

In some cases, one region may have seldom driver trajectories due to the limited number of drivers and the periodicity of drivers' working hours. We thus cannot compute the corresponding traffic. To solve this sparsity issue, we apply context-aware matrix factorization [23] to infer the traffic. The basic idea of context-aware matrix factorization is to factorize the spare matrix of traffic and the matrix of static urban features (i.e., the POI features and road network features) together. Therefore, the missing traffic values ($\bar{v}$ and σ_v) of a region are inferred by considering not only the traffic in its neighbor regions but also the traffic in the similar regions w.r.t. the static urban features.

2) DRIVER DISTRIBUTION FEATURE

The distribution of drivers has a direct impact on some KPIs. For example, if the number of drivers is much smaller than that of orders in a region R, the unmet demand and average response time there will be very large. Given a time slot, we can compute the number of drivers in each region R based on the GPS trajectories of drivers. In addition, to capture the status of the drivers in R, we also compute the idle ratio of each driver d in R within the given time slot T. Assuming that the GPS trajectory of driver d in region R within time slot T is $Tr_d{=}(p_1, p_2, \cdots, p_n)$, the corresponding idle ratio I_d of driver d is

$$I_d = \frac{\sum_{p_i, p_j \in Tr_d \wedge j=i+1 \wedge p_i.s=idle \wedge p_j.s=idle} (t_j - t_i)}{\sum_{i=1}^{n-1} (t_{i+1} - t_i)} \tag{6}$$

Considering that the staying times of drivers in a region could be different, we use weighted average idle ratio to represent the status of drivers in a region. Concretely, assuming that all the drivers appear in region R within time slot T are D, the weighted average idle ratio $\bar{I}$ is computed by

$$\bar{I} = \frac{\sum_{d \in D} w_d \cdot I_d}{\sum_{d \in D} w_d} \tag{7}$$

where weight parameter w_d is computed by

$$w_d = \frac{\sum_{i=1}^{n-1} (t_{i+1} - t_i)}{|T|} \tag{8}$$

Intuitively, weight parameter w_d computes the ratio of time slot T that driver d is in region R. Accordingly, the weighted deviation σ_I of idle ratio is computed by

$$\sigma_I = \frac{\sqrt{\sum_{d \in D} (w_d)^2 (I_d - \bar{I})^2}}{\sum_{d \in D} w_d} \tag{9}$$

3) WEATHER FEATURE

In general, weather plays an important role in affecting the KPIs of on-demand transport services. For example, when the weather is bad, more people tend to take cars to avoid an inconvenient trip to the nearby bus stations or subway stations. Therefore, in the proposed unified framework, we also take into consideration the factor of weather. The weather data is downloaded from Weather Underground which updates weather information every half hour. Roughly, we group all weather conditions into three categories, i.e., *good weather* (clear, haze, partly cloudy, mostly cloudy and

[5] https://www.wunderground.com/

scattered clouds), *bad weather* (light rain, rain, light rain showers and light thunderstorms), and *very bad weather* (rain showers, heavy rain showers and thunderstorms). Therefore, a weather feature with three values (i.e., 0: good weather, 1: bad weather, and 2: very bad weather) is computed.

E. PREDICTION MODEL TRAINING

Table 2 summarizes all the extracted features that will be used for predicting KPIs in on-demand transport services. In addition to the three categories of features discussed above, we also consider some other features including the region size, the time slot of the day, the day of the week, and whether the day is a public holiday.

TABLE 2. Summary for the extracted features.

Categories	Features
Historical KPI features	average KPIs in history, the KPIs in previous three time slots
Static urban features	POI (total POIs, number of POIs in each of the top-10 categories), road network (total length, lengths of five types of roads),
Dynamic urban features	traffic (average speed, speed deviation), driver (number of drivers, average idle ratio, idle ratio deviation), weather (good, bad or very bad)
Other features	region size, time slot of the day, day of the week, whether public holiday

With the extracted features, we employ *Least Absolute Shrinkage and Selection Operator* (Lasso) and *Gradient Boosting Regressor* (GBR) to train the prediction models. In general, Lasso performs better than other linear regression models because it can conduct feature selection during the training process. GBR also usually performs better than other ensemble models such as Random Forest Regression Model because it introduces boosting to improve the prediction performance. The two models (i.e., Lasso and GBR) are briefly introduced as below.

1) LASSO

Lasso [24] is a widely used linear regression model that enhances the prediction performance by conducting both variable selection and regularization. Given feature matrix X and label vector y, the objective function of Lasso is

$$\min_{w} \frac{1}{2n}||Xw-y||_2^2+\beta||w||_1 \tag{10}$$

where n is number of samples in feature matrix X; $||*||_2^2$ and $||*||_1$ computes the l_2-norm and l_1-norm, respectively; and parameter β is the weight parameter for regularization item $||w||_1$.

2) GBR

Gradient Boosting Regressor is an ensemble model that combines a set of weak prediction models to improve the accuracy of prediction. Concretely, GBR builds an additive model in a stage-wise fashion. Initially, a regression tree F_0 is fit on the training samples X and the corresponding residual is $F_0(X)-y$. A new regression tree F_1 is then fit on the negative gradient of loss function $\frac{1}{2}||F_0(X)-y||_2^2$. After repeating this fit operation M, a hyper-parameter, times, we can obtain a satisfactory prediction model.

In addition, we also compare our solution with Time-Varying Poisson Model [9] which is widely used for predicting demand in transport services, especially for the demand in traditional taxi services.

3) TRAIN AND TEST DATA

For each type of KPI, we have an order dataset of two months that contains $24*61*140=204,960$ samples in total, where 24, 61 and 140 correspond to the number of time slots (i.e., hours) in a single day, the number of days within two months, and the number of regions in Hong Kong. Each sample has all the features in Table 2 and has the corresponding KPI value in the next time slot as its label. We use the first 47 days (157,920 samples) for training while the last 14 days (47,040 samples) for testing.

4) TIME-DEPENDENT TRAINING

KPIs in on-demand transport services are dynamically changing over time. It is therefore difficult to train an efficient model for the whole day because the super-parameters in prediction models could be quite different for different time periods. To deal with this issue, we divide the time slots of a whole data into four groups with respect to the specific KPI. Concretely, for each type of KPI, we sort all the time slots according to the KPI value in descending order and select the first 25% time slots as data group G1, the second 25% as data group G2, the third 25% as data group G3, and the last 25% as data group G4. Then, we train a model and tune super-parameters separately for each data group, thus solving the issue of time-dependence.

5) FEATURE SELECTION

To reduce the complexity of trained models, we also conduct feature selection by using Lasso which is widely used for feature selection. For example, Figure 9 presents the top-10 most important features for predicting demand in data group G1, where the red vertical line corresponds to the optimal parameter β in Lasso. The features with weights obviously larger than zero around β are selected. Feature selection is also conducted on each group of data samples for other KPIs.

6) SUPER-PARAMETER TUNING

Among the three models (Time-varying Poisson, Lasso and Gradient Boosting Regressor), Time-varying Poisson does not have any super-parameters; Lasso has a weight parameter β for the item of regularization; Gradient Boosting Regressor has two major super-parameters, the number of estimators/iterations and the minimum number of samples in leaf nodes. For each type KPI, all these parameters are tuned using grid search for each group of data samples.

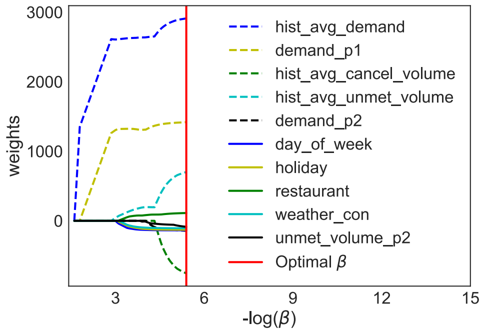


FIGURE 9. Feature selection for demand prediction in data group G1 using Lasso, where the meanings of top-10 most important features are the average demand in history, the demand in the previous time slot, the average cancelled demand in history, the average unmet demand in history, the demand in the past second time slot, day of the week, whether the day is a holiday, the number of restaurants, weather condition, and the unmet demand in the past second time slot, respectively.

TABLE 3. Experimental results for predicting KPIs of on-demand transport services, where ∗ indicates inapplicability.

KPI	Model	MAE	SMAPE	MAE-S	SMAPE-S
Demand	TVP	1.454	0.388	∗	∗
	Lasso	1.759	0.583	1.692	0.572
	GBR	**1.357**	**0.244**	**1.349**	**0.236**
Unmet demand	TVP	0.320	0.278	∗	∗
	Lasso	0.319	0.239	0.315	0.257
	GBR	**0.272**	**0.113**	**0.275**	**0.112**
Unmet rate	TVP	0.032	0.277	∗	∗
	Lasso	0.032	0.333	0.031	0.325
	GBR	**0.024**	**0.111**	**0.024**	**0.114**
Cancelled demand	TVP	0.551	0.408	∗	∗
	Lasso	0.547	0.520	0.543	0.523
	GBR	**0.487**	**0.185**	**0.476**	**0.181**
Cancel rate	TVP	0.086	0.403	∗	∗
	Lasso	0.084	0.692	0.082	0.667
	GBR	**0.061**	**0.184**	**0.061**	**0.188**
Avg. resp. time	TVP	40.299	0.513	∗	∗
	Lasso	38.817	0.731	38.385	0.730
	GBR	**32.179**	**0.348**	**32.227**	**0.346**

V. EXPERIMENTAL EVALUATION

A. EVALUATION METRICS

Both mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) are introduced as metrics to evaluate the performance of the proposed solution, where MAE is used for quantifying the absolute error while SMAPE is used for computing the percentage error. The meanings of MAE and SMAPE are briefly discussed as below.

B. MEAN ABSOLUTE ERROR (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \tag{11}$$

where $\hat{x}_i$ is the predicted value of real value x_i, and N is the number of evaluated samples.

C. SYMMETRIC MEAN ABSOLUTE PERCENTAGE ERROR (SMAPE)

$$SMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|u_i - \hat{u}_i|}{|u_i| + |\hat{u}_i|} \tag{12}$$

According to Eq. (12), SMAPE is between 0 and 1.

D. EXPERIMENTAL RESULTS

Table 3 presents the experimental results while using Time-varying Poisson (TVP), Lasso and Gradient Boosting Regressor (GBR), where MAE-S and SMAPE-S indicate that feature selection is conducted before training the model (inapplicable for TVP). In addition, the best results are highlighted in bold.

According to Table 3, GBR achieves the best performance in predicting all types of KPIs. TVP and Lasso have the similar performance in most cases except for some KPIs of high regularity. For example, TVP has a better performance in predicting demand than Lasso because demand is quite regular (cf. Figure 2). In addition, as suggested by Table 3, GBR achieves comparable performance after feature selection while the training time could be reduced greatly due to the smaller number of features. Lasso has the similar behavior as well.

As a case study, Figure 10 plots the predicted demand in a specific region using different methods, where the blue dashed line corresponds to the real demand and the orange solid line represents the predicted demand. According to the plots, the curve generated by Gradient Booting Regressor fits the real demand curve best.

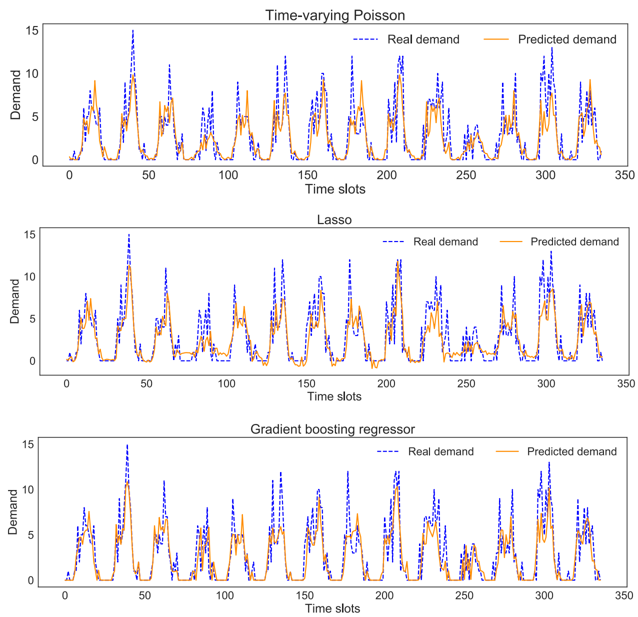


FIGURE 10. A case study for predicting the demand of a selected region using different prediction models.

E. ROBUSTNESS EVALUATION

To demonstrate the robustness of our solution, we also conduct prediction for two particular days, one public holiday (Day A in Figure 11) and one day with typhoon

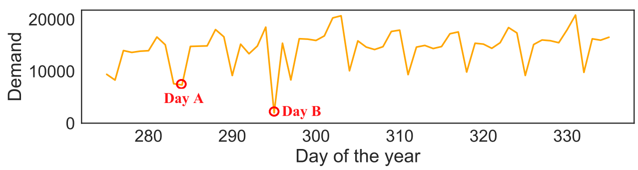


FIGURE 11. The daily demand in Hong Kong.

TABLE 4. Experimental results on Day A, a public holiday, where ∗ indicates inapplicability.

KPI	Model	MAE	SMAPE	MAE-S	SMAPE-S
Demand	TVP	4.564	0.487	∗	∗
	GBR	**1.286**	**0.310**	1.228	0.263
Unmet demand	TVP	0.154	0.092	∗	∗
	GBR	**0.045**	**0.040**	0.040	0.030
Unmet rate	TVP	0.017	0.055	∗	∗
	GBR	**0.006**	**0.020**	0.006	0.020
Cancelled demand	TVP	0.557	0.273	∗	∗
	GBR	**0.238**	**0.174**	0.237	0.166
Cancel rate	TVP	0.090	0.222	∗	∗
	GBR	**0.049**	**0.097**	0.049	0.098
Avg. resp. time	TVP	39.262	0.509	∗	∗
	GBR	**21.057**	**0.378**	20.921	0.403

TABLE 5. Experimental results on Day B, a day with bad weather, where ∗ indicates inapplicability.

KPI	Model	MAE	SMAPE	MAE-S	SMAPE-S
Demand	TVP	4.029	0.539	∗	∗
	GBR	1.951	0.423	1.832	0.422
Unmet demand	TVP	0.159	0.165	∗	∗
	GBR	**0.030**	**0.023**	0.030	0.023
Unmet rate	TVP	0.016	0.056	∗	∗
	GBR	**0.008**	**0.022**	0.008	0.022
Cancelled demand	TVP	0.516	0.377	∗	∗
	GBR	**0.286**	**0.197**	0.262	0.209
Cancel rate	TVP	0.086	0.257	∗	∗
	GBR	**0.101**	**0.143**	0.100	0.143
Avg. resp. time	TVP	36.112	0.561	∗	∗
	GBR	**20.508**	**0.483**	19.004	0.479

(Day B in Figure 11). According to Figure 11, the demands on the two selected days are greatly different than usual.

Tables 4 and 5 show the results of prediction when using Time-varying Poisson model and GBR model. Since GBR performs better than Lasso, we thus omit the results of Lasso to save space. According to the two tables, GBR still achieves good prediction for different KPIs and performs much better than Time-varying Poisson model. Therefore, our solution has good robustness to predict KPIs in various scenarios.

VI. CONCLUSION

In this work, we propose a unified framework to predict multiple KPIs for on-demand transport services. The proposed framework not only considers the temporal regularity of KPIs, but also adapts well to the dynamic changes of urban contexts. More importantly, it fuses data from multiple domains to obtain comprehensive knowledge about the KPIs, thus achieving a satisfactory performance in the prediction.

However, our framework still has some limits. First, it cannot adapt quickly to predict some outliers of KPIs if there are no similar historical cases. For example, when a serious traffic accident happens in some region, the KPIs there will be seriously affected. If three are no similar traffic accidents happened in history, it will be difficult to achieve accurate prediction for KPIs in this case. Second, the proposed framework only reports the KPIs in the next time slot, and do not provide solutions or operation recommendations to avoid bad KPIs.

As for the future work, considering the limits of our framework, it would be of high importance to detect the outliers of KPIs since they seriously affect the service quality of on-demand services yet are difficult to predict using the general prediction models. Another promising study is to devise effective solutions to improve the KPIs of on-demand services, e.g., behavior analysis for drivers and passengers, passenger-driver relationship analysis, dynamic pricing, and intelligent dispatching.

REFERENCES

[1] P. S. Rossi, D. Ciuonzo, and T. Ekman, "HMM-based decision fusion in wireless sensor networks with noncoherent multiple access," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 871–874, May 2015.

[2] D. Ciuonzo and P. S. Rossi, "Distributed detection of a non-cooperative target via generalized locally-optimum approaches," *Inf. Fusion*, vol. 36, pp. 261–274, Jul. 2017.

[3] J. Chen, K. H. Low, Y. Yao, and P. Jaillet, "Gaussian process decentralized data fusion and active sensing for spatiotemporal traffic modeling and prediction in mobility-on-demand systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 3, pp. 901–921, Jul. 2015.

[4] L. Moreira-Matias, J. Gama, M. Ferreira, and L. Damas, "A predictive model for the passenger demand on a taxi network," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2012, pp. 1014–1019.

[5] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, 2013.

[6] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "On predicting the taxi-passenger demand: A real-time approach," in *Proc. Prog. 16th Portuguese Conf. Artif. Intell. (EPIA)*, 2013, pp. 54–65.

[7] X. Li *et al.*, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[8] W. Jiang, T. Wo, M. Zhang, R. Yang, and J. Xu, "A method for private car transportation dispatching based on a passenger demand model," in *Proc. 2nd Int. Conf. Internet Vehicles-Safe Intell. Mobility (IOV)*, 2015, pp. 37–48.

[9] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Philadelphia, PA, USA, Aug. 2006, pp. 207–216.

[10] D. Liu, S. Cheng, and Y. Yang, "Density peaks clustering approach for discovering demand hot spots in city-scale taxi fleet dataset," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2015, pp. 1831–1836.

[11] H. Chang, Y.-C. Tai, and J. Y.-J. Hsu, "Context-aware taxi demand hotspots prediction," *Int. J. Bus. Intell. Data Mining*, vol. 5, no. 1, pp. 3–18, 2010.

[12] K. Zhang, Z. Feng, S. Chen, K. Huang, and G. Wang, "A framework for passengers demand prediction and recommendation," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jun./Jul. 2016, pp. 340–347.

[13] B. Jäger, M. Wittmann, and M. Lienkamp, "Analyzing and modeling a city's spatiotemporal taxi supply and demand: A case study for munich," *J. Traffic Logistics Eng.*, vol. 4, no. 2, pp. 147–153, 2016.

[14] K. Zhao, D. Khryashchev, J. Freire, C. Silva, and H. Vo, "Predicting taxi demand at high spatial resolution: Approaching the limit of predictability," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2016, pp. 833–842.

[15] N. Mukai and N. Yoden, "Taxi demand forecasting based on taxi probe data by neural network," in *Proc. 5th Int. Conf. Intell. Interact. Multimedia Syst. Services (IIMSS)*, 2012, pp. 589–597.

[16] A. Anwar, M. Volkov, and D. Rus, "ChangiNOW: A mobile application for efficient taxi allocation at airports," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 694–701.

[17] J. Bischoff, M. Maciejewski, and A. Sohr, "Analysis of Berlin's taxi services by exploring GPS traces," in *Proc. Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, 2015, pp. 209–215.
[18] J. Lee, I. Shin, and G.-L. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," in *Proc. 4th Int. Conf. Networked Comput. Adv. Inf. Manage.*, 2008, pp. 199–204.
[19] A. Afian, A. Odoni, and D. Rus, "Inferring unmet demand from taxi probe data," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2015, pp. 861–868.
[20] D. Shao, W. Wu, S. Xiang, and Y. Lu, "Estimating taxi demand-supply level using taxi trajectory data stream," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 407–413.
[21] Y. Lu, S. Xiang, and W. Wu, "Taxi queue, passenger queue or no queue?" in *Proc. Int. Conf. Extending Database Technol. (EDBT)*, 2015, pp. 593–604.
[22] Y. Huang and J. W. Powell, "Detecting regions of disequilibrium in taxi services under uncertainty," in *Proc. SIGSPATIAL*, 2012, pp. 139–148.
[23] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2014, pp. 1027–1036.
[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

JIHONG GUAN received the bachelor's degree from Huazhong Normal University in 1991, the master's degree from the Wuhan Technical University of Surveying and Mapping (merged into Wuhan University since 2000) in 1991, and the Ph.D. degree from Wuhan University in 2002.

She was an Assistant Professor with the Department of Computer, Wuhan Technical University of Surveying and Mapping, from 1991 to 1997, where she has been an Associate Professor since 2000. She was an Associate Professor with the School of Computer, Wuhan University, from 2000 to 2003, where he has been a Professor since 2003. She is currently a Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. Her research interests include databases, data mining, distributed computing, bioinformatics, and geographic information systems.

WEILI WANG received the B.Sc. and M.Sc. degrees in computer science from Nanchang University, China, in 2006 and 2009, respectively.

He is currently pursuing the Ph.D. degree with the Department of Computer Science and Technology, Tongji University, China. He is also with the Department of Computer, Nanchang University. His research interests include spatial data management and query processing, distributed computing, and geographic information systems.

WENGEN LI received the B.E. and Ph.D. degrees in computer science from Tongji University, Shanghai, China, in 2011 and 2017, respectively, and the dual Ph.D. degree in computer science from The Hong Kong Polytechnic University in 2018.

He is currently a Post-Doctoral Fellow with the Department of Computing, The Hong Kong Polytechnic University. His research interests include spatial data management and big data analytics for human mobility and urban logistics. He is a member of the China Computer Federation.

SHUIGENG ZHOU (M'07) received the bachelor's degree from the Huazhong University of Science and Technology, in 1988, the master's degree from the University of Electronic Science and Technology of China, in 1991, and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2000.

He was an Engineer with the Shanghai Academy of Spaceflight Technology from 1991 to 1997, where he has been a Senior Engineer since 1995. He was a Post-Doctoral Researcher with the State Key Laboratory of Software Engineering, Wuhan University, from 2000 to 2002. He is currently a Professor with the School of Computer Science, Fudan University. His research interests include data management, data mining, and bioinformatics.