DETECTING AND ANALYZING URBAN REGIONS WITH HIGH IMPACT OF WEATHER CHANGE ON TRANSPORT

[1]Mr.D.Manoj, [2]Mr.P.A.Selvaraj
[1]Student, [2]Assistant professor
Computer Applications, Kongu Engineering College, Perundurai Erode, Tamilnadu , India

Abstract : ***We focus on two fundamental questions that are unprecedentedly important to urban planners to understand the functional characteristics of various urban regions throughout a city, namely, (i) how to identify regional weather-traffic sensitivity index throughout a city, that indicates the degree to which the region traffic in a city is impacted by weather changes; (ii) among complex regional features, such as road structure and population density, how to dissect the most influential regional features that drive the urban region traffic to be more vulnerable to weather changes. However, these two questions are nontrivial to answer, because urban traffic changes dynamically over time and is essentially affected by many other factors, which may dominate the overall impact. We make the first study on these questions, by developing a weather-traffic index (WTI) system. The system includes two main components: weather-traffic index establishment and key factor analysis. Using the proposed system, we conducted comprehensive empirical study in Shanghai, and the weather-traffic indices extracted have been validated to be surprisingly consistent with real world observations. Further regional key factor analysis yields interesting results***

Index Terms - Trajectory analysis, weather-traffic index, traffic prediction, urban computing.

1.INTRODUCTION

URBAN computing connects urban sensing, data management, data analytic and service providing into a recurrent process for an unobtrusive and continuous improvement of people's lives, city operation systems and the environment . The aim is to solve a variety of emerging city problems, such as traffic congestion, energy consumption, and pollution, based on the data of traffic flow, human mobility, and geographical data, etc. In particular, many works have been done to investigate the impact of inclement weather to traffic.

The decreasing temperature in very cold days will freeze the roads and influence the transport performance, etc. Table 1 describes the general relevance of the impact of weather change to transport in US. In July 21st, 2012, Beijing faces its largest rainstorm since 1951, with an average rainfall of 164 millimeters. According to the news report [7], there are 77 people died in this catastrophic natural disaster. The transport of Beijing suffered from various contingencies due to the serious flood, as shown in Fig. 1. During that time, a variety of photos titled "see the sea in Beijing" widespread on the Internet.

This disaster not only shows the serious problems of the urban transport system of Beijing, but also inspires our research interest: how can we identify those regions being highly influenced by weather change on transport? The early works often focus on the correlation of weather and traffic in some particular roads where devices have been deployed to continuously collect traffic data. By analyzing the traffic change in different weather conditions, the traffic prediction can be better preformed considering the weather forecast. However, the weather-traffic correlation covering most roads throughout a city (known as regional weather-traffic sensitivity index or for simplicity weather-traffic index (WTI)) is still an open problem vain in spite of the practical value in our daily life.

One essential reason is the lacking of effective traffic monitoring system in city-wide scale. Another open problem is how to disclose the key factors behind the weather-traffic index, to explain the reason why some regions in a city are more vulnerable to inclement weather and others are not. These factors are the regional features including the density of roads, the number of road intersections, the number of points of interest (POIs), the traffic volume, the average age of the household, the density of buildings and more in the surrounding regions.

The weather-traffic index throughout a city and the knowledge of key factors behind the correlation provides effective support to help government agent to understand the functional character of districts throughout a city, to improve traffic performance and to learn the key factors in urban planning,

Fig. 1. The rainstorm of Beijing in the year of 2012.

TABLE 1
Changes in Climate and Weather Relevant on US Transport [6]

Change in Climate or Weather	Likelihood
Decreases in very cold days	Virtually certain
Increases in Arctic temperatures	Virtually certain
Later onset of seasonal freeze and earlier onset of seasonal thaw	Virtually certain
Sea level rise	Virtually certain
Increases in very hot days and heat waves	Very likely
Increase in intense precipitation events	Very likely
Increases in drought conditions for some regions	Likely
Changes in seasonal precipitation and flooding patterns	Likely
Increases in hurricane intensity	Likely
Increased intensity of cold-season storms, with increases in winds and in waves and storm surges	Likely

2. RELATED WORK

Urban computing works often focus on a particular city problem, such as traffic congestion, energy consumption, and pollution, based on the data of traffic flow, human mobility, and geographical data, etc.

The real-time and fine-grained air quality information throughout a city, based on the air quality data reported by existing monitor stations and a variety of data sources observed in the city. In [9], they tried to identify the hot spots of moving vehicles in an urban area via a novel, non-density-based approach, called mobility-based clustering. In [10], they proposed a framework, called DRoF, to discover regions of different functions in a city using both human mobility among regions and points of interests located in a region. In [11], the authors tried to sense the refueling behavior and citywide petrol consumption in realtime, based on the trajectories of vehicles. In [12] and [13], they tried to discover the traveling companions and gathering patterns of vehicles, respectively.

Being an important topic in urban computing and cross domain data analytics [14], the early research on the relation between weather and traffic is mainly based on quantitative analysis and statistical methods. For example, in [3], they presented an algorithm for forecasting physical road surface conditions based on weather and road surface data they have collected, and aim to identify icy roads during a cold weather in advance in order to predict the impact to traffic. In [4],

They proposed a crash-likelihood prediction model based on both real-time traffic flow variables measured through series of underground sensors and the rain data collected at weather stations in order to alarm potential crash occurrence in advance. In [5], they developed a neurowavelet prediction algorithm to forecast hourly traffic flow considering the effect of rainfall. The experiments show that the rainfall data successfully augments the traffic flow data as an exogenous variable in periods of inclement weather. The early works focus on some particular roads where devices have been deployed to continuously collect traffic data. None of them investigated the weather-traffic correlation throughout a city and conduct analysis of the key factors behind the regions whose traffics are highly influenced by inclement weather

3.OVERVIEW

This work aims to develop a weather-traffic index system which performs two tasks: establishment of weather-traffic index throughout a city and analysis of key factors behind the index. The framework of the proposed system is shown in Fig. 2, which consists of three functional components:

- **Data preparation**: The road networks in the city of interest is partitioned into cells via Voronoi diagram where the seeds are road intersections. For each cell, the traffic parameters are extracted from taxi trajectories and the regional features are collected. The weather information of the same period of time is also collected. The details are presented in Section 4.
- **Weather-traffic index establishment**: The weather-traffic index is established for each cell by analyzing traffic data and weather data. In specific, given a cell g, the weather-traffic index rðgÞ is a value indicating the extent to which the traffic parameter in g is affected by weather. This component is discussed in Section 5.
- **Factor analysis**: The input includes the established weather-traffic index and the regional features. The aim is to identify which regional features make traffic in cells vulnerable to inclement weather. In particular, the weights of regional features are quantitatively measured

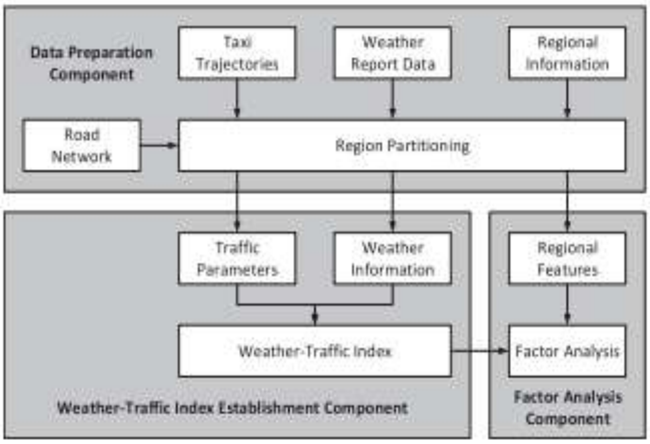


Fig. 2. Framework of weather-traffic index (WTI) system.

4.DATA PREPARATION

In this section, we introduce the data preparation component which partitions the city into fairly distributed regions, and collects relevant source data for each region.

4.1 Region Partitioning

A straightforward region partitioning method is region-oriented partitioning such as in [8] where the city region is split into equal size grids. However, this partitioning method is improper if the traffics of road networks in grids are concerned. The reason is that the road networks in a city are often distributed unevenly. For example, the road networks are typically much denser in the urban areas than that in the rural areas.

As a consequence, the road networks in some grids are highly dense and in some grids are highly sparse. This situation motivates us to apply a different region partitioning method. Our method is to partition the city region using Voronoi diagram [15]. A Voronoi diagram is a partitioning of a plane into regions (or cells) based on the distance to points (or seeds) in a specific subset of the plane, and the shapes and sizes of the cells differ from each other. In this paper, we choose road intersections as the seeds. We call such partitioning method as road-intersection-oriented partitioning. In particular, if several road intersections are very close to each other, for example within 50 meters, they are grouped together as a complex intersection.

So, each cell includes at least one road intersection and the road segments connected with this intersection. The indices of all cells are obtained following the equal procedure no matter they are in dense and non-dense areas. The road-intersection-oriented partitions in Shanghai is shown in Fig. 3 where the seeds are the intersections of major roads. We observe that the cells are relative small in the urban areas while the cell tends to be large in rural areas. This partitioning method has two desirable properties. The first is the relatively even distribution of road networks in all cells.

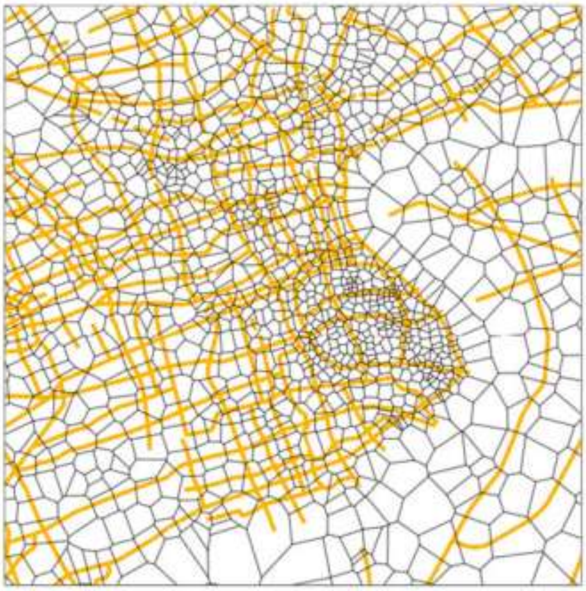


Fig. 3. The Voronoi diagrams partitions in Shanghai. The under layer represents the road networks.

TABLE 2
Specifications of Weather Report Data

Attribute	Description
Time	Time of the weather report.
Temperature	Temperature in Celsius degrees.
Dew Point	The temperature at which the air must be cooled for water vapor to condense, forming water droplets, fog, or clouds.
Humidity	The relative amount of water vapor in the air.
Wind Speed	Speed of wind shown in km/h.
Wind Gust	The maximum wind speed in km/h.
Wind Direction	The direction of wind in degrees.
Visibility	The ability to see an object in the atmosphere in km.
Pressure	The Atmospheric air pressure in millibars.
Wind Chill	The perceived decrease in air temperature felt by the body on exposed skin due to the flow of air.
Heat Index	An index that combines air temperature and relative humidity to estimate the human-perceived equivalent temperature.
Precipitation	The condensation in mm of atmospheric water vapor that falls under gravity, including drizzle, rain, sleet, snow, etc.
Condition	Weather condition, e.g., clear, rainy, and cloudy.

4.2 Source Data

The input of the system includes the road networks, traffic data, and regional features in the city of interest, and the weather data in the same period of time. A road network GðV; EÞ consists of a set of road segments E and a set of road intersections V . A road segment in E is associated with its type, length, speed limit, two end points and other meta information. A road intersection in V is associated with its location (i.e., latitude and longitude) and type.

The carriageway between two road intersections in E may consist of multiple edges in E connected in sequence. From traffic data, a certain traffic parameter of interest, such as average speed, can be extracted. Traffic parameter can be classified in terms of one of the following: quantity measures, e.g., “how much or at what rate is traffic moving or waiting to move?”; quality assessment measures, e.g., “how well is traffic moving?”; movement measures, e.g., “where is traffic coming from and going to?”; and composition / classification measures, e.g., “what kind of traffic is moving?”.

While all kinds of traffic parameters can be applied in our weather-traffic index, this work use average speed as an example. Speed expresses the rate at which traffic is moving and, therefore, is a natural measure of the quality of the flow. In this work, time mean speed (also called average speed) is used as the traffic parameter, which is defined as the arithmetic mean of individual spot speeds that are recorded over a selected time period. An adequately sized sample of spot speeds is needed to ensure that the time mean speed approximates the population mean to within the desired accuracy. The traffic parameters are extracted from large volume of taxi trajectory data collected.

In our study, the average speed of the driving taxis in each cell are calculated. In particular, the average speed is split into seven classes: less than 10, 10-30, 30-50, 50-70, 70-90, 90-110 km/h, and more than 110 km/h. Since traffic parameters are categorical results and our objective is to establish index, we split the continuous variables because 1) it reduces the complexity of the problem, and 2) it well supports our objective. In particular, if continuous values are used, the main ideas proposed in this work are still applicable with trivial modification.

The average speed of one road segment is subject to the traffic parameter of that road segment only, which is not comparable with other road segments. For example, the average speed of 30 km/h reflects drastically different traffic condition on a small local street and a highway. Hence, in this paper, we only compare the changes of the average speed on each road segment separately. Weather is the state of the atmosphere, to the degree that it is hot or cold, wet or dry, calm or stormy, clear or cloudy. The details of the weather data used in this paper are described in Table 2 in Section 7.1. If a particular weather condition is interested, such as rain, the weather-traffic index can be specialized as rain-traffic index and accordingly the factor analysis is specialized to rain as well

5.WEATHER-TRAFFIC INDEX ESTABLISHMENT

The weather data and traffic data from data preparation component is the input of weather-traffic index establishment component. The intuition that the traffic is influenced by weather can be proven by the example shown in Fig. 4. It illustrates the average speeds in different cells in Shanghai at the same time slots in two different weather conditions: cloudy and rainy.

It is clear that the average speed in rainy days is generally lower than that in cloudy days. At the same time, it also demonstrates that the average speeds in some cells are unchanged in the rainy days and in cloudy days. Therefore, weathertraffic index is necessary to indicate the impact of weather to traffic in different cells.

Given a cell g, its value in weather-traffic index is the correlation between traffic and weather, denoted as rðgÞ. rðgÞ takes value from a discrete range, such as ½1; 2; 3; 4; 5. The following section will discuss how to detect such correlation

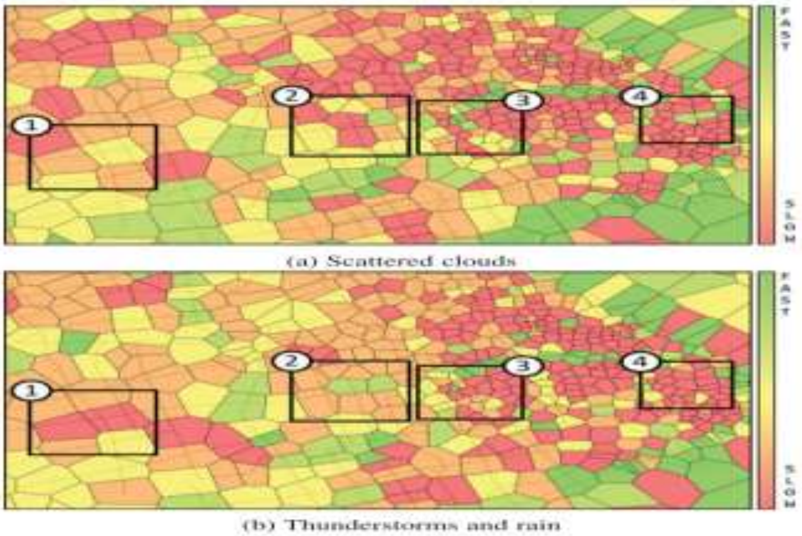


Fig. 4. The average speed at 14:00 on two days in summer in Shanghai, where the weather is scatter cloudy (top) and rainy (bottom).

5.1 Correlation Detection

In a cell, for detecting the correlation between the traffic speed, denoted as Ft, and weather, denoted as Fw, a simple method is to train a classifier which infers directly from Fw to Ft, as shown in Fig. 5. The input is the weather represented as a feature vector and the output is one of the seven speed classes. The trained classifier is tested. If the inference accuracy is high, it means the correlation between the traffic and weather is high in this cell; otherwise, the correlation is low. This method is commonly used in statistics to measure the correlation between two random variables.

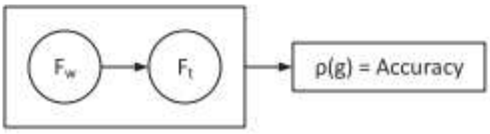


Fig. 5. A simple weather-traffic correlation detection method where traffic speed is directly inferred from weather.

However, we observed critical weakness of this method in correlation detection between traffic and weather. This is because there are many other reasons which impact traffic. For example, the traffic in peak-hour differs from that in non-peak hours, the traffic accident in one road segment will influence the traffic in nearby road networks, and the road works slow down the average speed, etc. Compared to weather, these reasons are dominant in most cases.

Therefore, the main challenge in weather-traffic index establishment is to separate the impact of weather to traffic in each cell from other reasons. To address this challenge, we propose a novel method inspired by the Granger causality test [16]. The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. A time series X is said to Granger-cause Y if it can be shown that those X values provide statistically significant information about future values of Y . Hence, we say that a variable X that evolves over time Granger-causes another evolving variable Y if predictions of the value of Y based on its own past values and on the past values of X are better than predictions of Y based only on its own past values. In this paper, the initiative is to train a traffic prediction model which considers all other reasons besides weather, and then train a traffic prediction model which considers all other reasons and weather.

We observe the difference between the inference accuracies of the two models. If the accuracy is improved after considering weather, it indicates that the weather does impact the traffic in this cell in general; otherwise, the impact of weather is uncertain in this cell. The overview of our method is shown in Fig. 6. The traffic prediction models are trained separately in different time slots.

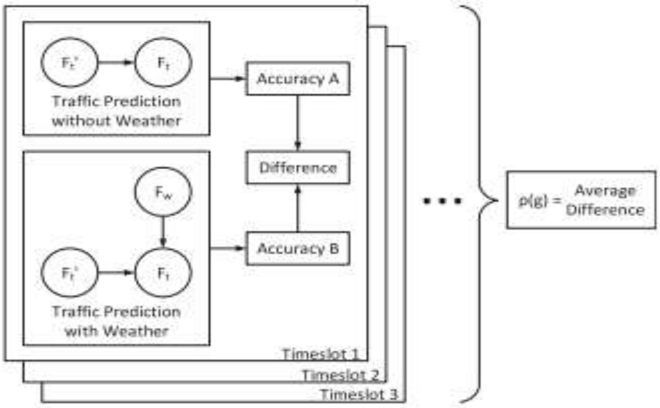


Fig. 6. The weather-traffic correlation detection method used in this paper.

5.2 Traffic Prediction

Traffic prediction is a well studied problem. Since early 1980s, univariate time series models, mainly Box-Jenkins Auto-Regressive Integrated Moving Average (ARIMA) [17] and Holt-Winters Exponential Smoothing (ES) [18], have been widely used in traffic prediction. In the last decade, neural network models have also been used in forecasting travel time [19]. In [20], spatial-temporal characteristics of traffic events are considered in training traffic prediction models. In [21], authors use AQ21, a natural induction system that learns and applies attributional rules, to predict traffic by autonomous agents within a vehicle route planning system. In [22], they estimate the traffic flow of a road segment by analyzing taxi trajectories. A recent study successfully uses the weather situations as supplementary information in traffic prediction model to enhance the prediction accuracy [4]. In this work, any traffic prediction model can be used.

6. FACTOR ANALYSIS

In this section, our discussion is based on the assumption that the weather-traffic indices of all cells have been certainly assigned. The weather-traffic index indicates which cells are correlated with weather in terms of traffic. It provides the possibility for us to investigate the key factors behind the correlation. The factors are the regional features, denoted as Fr, as shown in Table 4 in Section 7.1. The factor analysis identifies the key factors and their weights contributing to the weather-traffic indices of cells. In other words, it discloses what regional features make the traffic in some cells vulnerable to inclement weather.

6.1 Key Factor Verification by Index Inference (KFVII)

Given a set of regional features, our methodology verifies they are the key factors based on the following intuition. The weather-traffic index of one region can be inferred from the indices of its closely located (or adjacent) cells. The intuition is feasible because all the regions are connected by the road network, which can directly show the sensitivities of regions against weather. Based on the intuition, give a set of regional features, if the inference accuracy is satisfactory using For as input, it indicates that such set of regional features are the key factors. The intuition leads to the model as shown in Fig. 7. In Fig. 7, the parent node gu specifies the source cell, and the child node gi is a set of observed cells which are closely located to gu. This model is not symmetric since gi ! gu may have a different probability comparing with gu ! gi. The inference model can be any graphical classifier but we propose to use naive Bayes classifier [28], because the location closeness can be naturally considered by naive Bayes classifier. Since different cells have different numbers of neighboring cells, it is hard to use other classifiers such as logistic regression, SVM, neural network, and random forest where the number of input features is fixed. In this situation, Naive Bayes classifier is a reasonable choice. The following sections describe the details of the Naive Bayes classifier, from constructing the marginal distribution to the detailed index-index inference method.

6.1.1 Marginal Distribution

The marginal distribution used in this paper is shown in Fig. 8. A marginal distribution describes the probability distribution of the regions contained in a similarity subset [28]. Specifically in this paper, it describes the probability of one region being the index of i given one of its adjacent regions with index j, if the two regions have a certain similarity. The similarities are split into subsets because the probability distributions may vary upon different similarities. In this paper, we use cosine similarity in terms of regional features

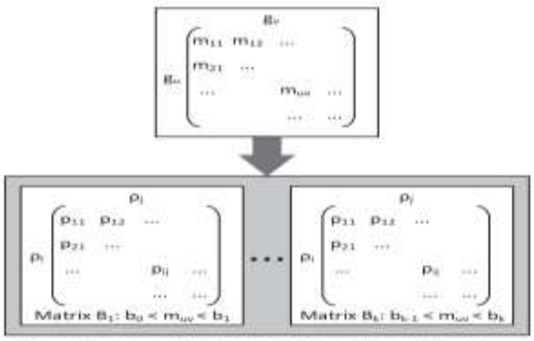


Fig. 8. Converting from similarity matrix to marginal distribution.

as shown in Equation (1) to describe the similarity m_{uv} between two regions g_u and g_v

$$m_{uv} = \frac{F_r^u \cdot F_r^v}{\|F_r^u\| \|F_r^v\|} = \frac{\sum_{i=1}^{n} F_r^u(i) \times F_r^v(i)}{\sqrt{\sum_{i=1}^{n} (F_r^u(i))^2} \times \sqrt{\sum_{i=1}^{n} (F_r^v(i))^2}}. \quad (1)$$

According to the similarity of regional features, all pairs of adjacent cells are clustered into k groups. Suppose b_0 is the minimum similarity and b_k is the maximum similarity. The similarity ranges of the k groups are $[b_0, b_1], \ldots, [b_{k-1}, b_k]$ as shown in Fig. 8. The group of $[b_{i-1}, b_i]$ only contains the pairs whose similarities are in between b_{i-1} and b_i. So, the pairs in the same group have the similar similarity. In the group of $[b_{i-1}, b_i]$, the pairs of cells are summarized to marginal distribution matrix B_i. The rows of B_i are the weather-traffic indices of g_u and the columns of B_i are weather-traffic indices of g_v. Specifically, when the weather-traffic index of g_u is ρ_i, the probability that the weather-traffic index of g_v is ρ_j is recorded in p_{ij}. For example, there are 500 pairs of cells in group of $[b_{i-1}, b_i]$. Suppose, in 200 pairs of them, one cell has index 2 and the number of another cells whose index is 1 is 50. Then p_{21} in matrix B_i is 0.4. It indicates that, if two cells have similarity in terms of regional features in between b_{i-1} and b_i, and the weather-traffic index of one cell is 1, the probability that the weather-traffic index of the other cell is 2 is 0.4. Formally,

$$p_{ij} = Pr(\rho(g_u) = \rho_i | \rho(g_v) = \rho_j) = \frac{|\rho(g_u) = \rho_i, \rho(g_v) = \rho_j|}{|\rho(g_v) = \rho_j|}. \tag{2}$$

$$\begin{aligned} & Pr(\rho(g_u) = \rho_u | \rho(g_1) = \rho_1, \rho(g_2) = \rho_2, \ldots) \\ &= \frac{Pr(\rho(g_1) = \rho_1, \ldots | \rho(g_u) = \rho_u) * Pr(\rho(g_u) = \rho_u)}{\sum_{i=1}^{k} Pr(\rho(g_1) = \rho_1, \ldots | \rho(g_u) = \rho_i) * Pr(\rho(g_u) = \rho_i)} \\ &= \frac{Pr(\rho(g_1) = \rho_1 | \rho(g_u) = \rho_u) * \cdots * Pr(\rho(g_u) = \rho_u)}{\sum_{i=1}^{k} Pr(\rho(g_1) = \rho_1 | \rho(g_u) = \rho_i) * \cdots * Pr(\rho(g_u) = \rho_i)}. \end{aligned} \tag{3}$$

Given a cell g_u, the marginal distribution allows naïve Bayes classifier to infer which value the weather-traffic index of g_u is most likely to be, based on the weather-traffic indices of its adjacent cells $\rho(g_1)$, $\rho(g_2)$, ... The inference accuracy of 10-fold cross validation is then obtained.

VI.c) Weight Estimation of Regional Features

Given a set of regional features, some of them may have trivial impact to weather-traffic index, or are just noise. This requires us to test the weight of each regional feature through a feature selection [29] process. There are many feature selection methods could be used in this paper. For example, Fisher score [30], where features are scored by considering that features with high quality should assign similar values to instances in the same class and different values to instances from different classes; and ReliefF [31], [32], which selects features to separate instance from different classes. In this paper, we uses the following method similar as mutual information based methods [33], [34], [35].

Suppose a regional feature has nontrivial impact to weather-traffic index. Let us remove this regional feature from the set of regional features. We can use the KFVII method in Section 6.1 to test whether the remaining set of regional features is still the set of key factors which results in high overall accuracy. If not, it is a strong signal that the removed regional feature is very important; otherwise, it is less important. We use dðFi r Þ to denote the weight of the regional feature Fi r . Look closely, the similarity of every two adjacent cells are recomputed in terms of the remained regional features, as well as the marginal distribution. If the inference accuracy is increased more, the removed regional feature has more weight.

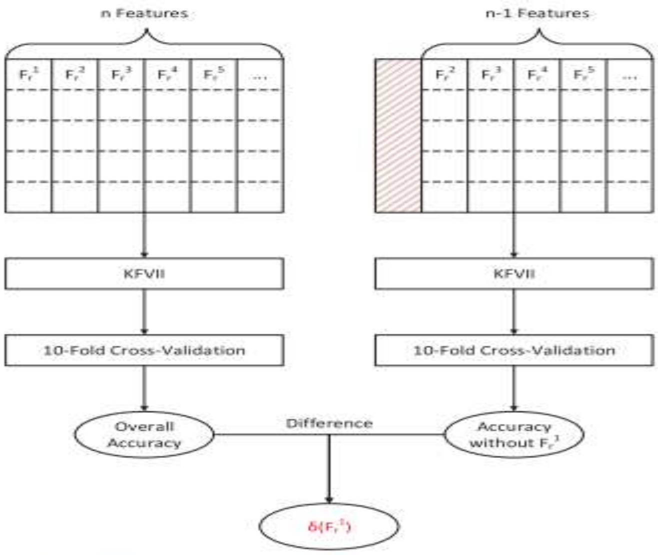


Fig. 9. Weight estimation of regional feature F_r^1.

7. EMPIRICAL STUDY

In this section, we conduct empirical study using the proposed weather-traffic index system in Shanghai. We first introduce our data sources

7.1 Datasets

The input of our weather-traffic index system includes
(i) road network data of Shanghai,
(ii) taxi trajectory data collected in Shanghai,
(iii) weather report data of the same period of time and
(iv) regional information data.

7.1.1 Road Network Data

The road network data of Shanghai is provided by the government, where a road (or precisely a road segment) is defined as the carriageway between two intersections. An expressway or a large avenue may have two different road Fig. 8. Converting from similarity matrix to marginal distribution. A road network is consisted of a set of roads. There are in total seven levels of roads: national expressway, city expressway, regular highway, large avenue, primary way, secondary way, and regular road [36]. We consider the first four levels of roads as major roads and the other three levels as minor roads. In this study, only the major roads are used in region partitioning and the minor roads are ignored.

7.1.2 Weather Report Data

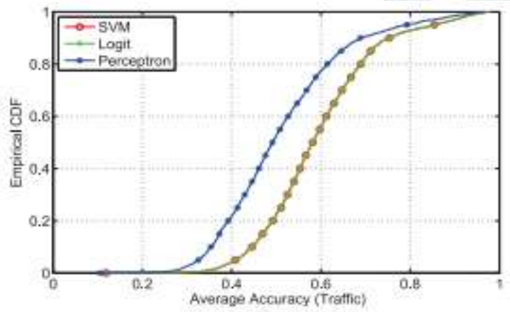
Weather report data are collected from Weather Underground (wunderground.com), which is a leading website on commercial weather service providing weather forecast and historical weather information. The weather data contain rich information covered by 14 weather features, including temperature, wind speed, precipitation, etc. In Table 2, we summarize all 14 weather features used in this paper, and they are processed all together. For data alignment, the collected weather reports in Shanghai cover the same period of time as that of taxi trajectory data, i.e., January 2006 to November 2007. The weather is reported on hourly basis. Accordingly, we split day time into time slots by hours.

7.1.3 Trajectory Data

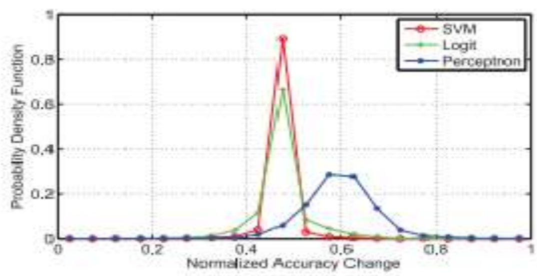
A trajectory is represented as a series of spatial-temporal points [37], where each point is associated with additional information including the driving speed. Our trajectory data of 115.2 GB are collected from 4,529 taxis in Shanghai from January 2006 to November 2007. The average sampling rate of the dataset is about 20 seconds. Table 3 lists the fields recorded in the trajectory data. By extracting the driving speeds of all taxis in each Voronoi cell at each time slot, the average speed is obtained

7.2 Weather-Traffic Index

Weather-traffic index in Shanghai is constructed using the weather-traffic index establishment method introduced in Section 5. Briefly, for each cell in each time slot, the average speed is inferred using traffic prediction model with/without weather. The inference accuracy difference indicates the sensitivity of this cell at this time slot. The average of the inference accuracy differences at all time slots indicate the sensitivity of the cell, depending on the fraction of the time that the cell was experiencing abnormal weather. In particular, the traffic prediction model without weather uses the average speeds of previous days of the same cell at the same time slot as the input. The purpose is to predict the current average speed with minimal weather impact since the previous days are in “good” and “bad” weather conditions and as a result the weather impact is trade-off. In the traffic prediction model with weather, the weather features of the current day is used as the additional input features in the traffic prediction.

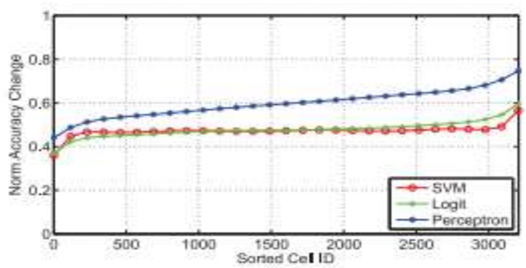


(a) The cumulative distribution of the accuracy of traffic prediction without weather.



(c) The probability distribution of the traffic prediction accuracy changes in all the regions with/without weather.

Fig. 10. Evaluation of weather-traffic index.



(b) The traffic prediction accuracy changes in all the regions with/without weather.

7.2.1 Validation

The effectiveness of weather-traffic index established have been verified against the observations in the real world. Fig. 11 shows the resulting weather-traffic index of the regions in the urban areas of Shanghai. In Fig. 11, a positive weather-traffic index describes a region with high impact of weather change on transport, and a negative weather-traffic index describes a region with low impact of weather change on transport. There are five regions with very high weathertraffic index as labeled 1-5 in the figure, and the details of these regions are shown in Table 5. In practical, there are often many people walking near a tourism attraction, and when there is a rain, the tourists may have rush home.

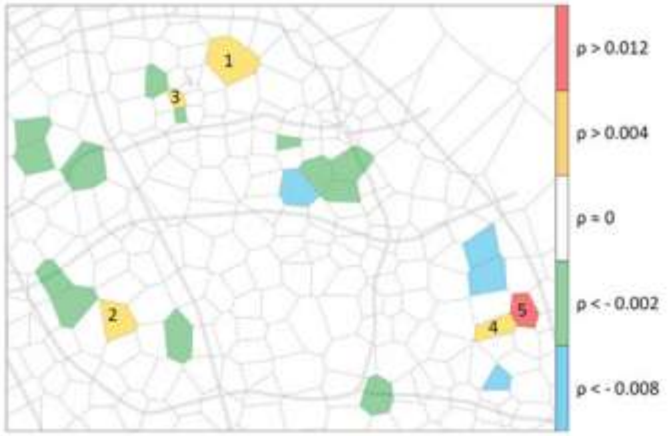


Fig. 11. The weather-traffic index of the regions in the urban areas of Shanghai. The details of the labeled regions are shown in Table 5.

TABLE 5
Details of Regions with High Weather-Traffic Index

ID	Description
1	Yu Garden, a tourism attraction.
2	Shanghai Confucian Temple, an old temple with many restaurants around.
3	Shanghai Town God's Temple, a tourism attraction.
4	An area with many old buildings and construction sites.
5	Construction sites (Bund House, a high-end residential community is built several years later).

7.4 Factor Analysis

Given any set of regional features, we verify they are key factors to weather-traffic index or not by KFVII introduced in Section 6.1, and then generalize the weight of the set of regional features.

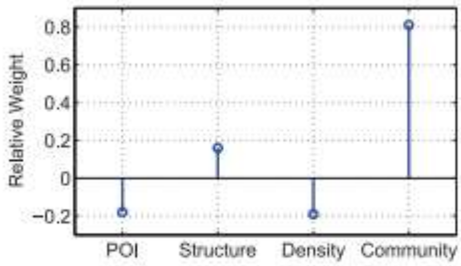


Fig. 15. The weights of the regional feature sets in Table 4.

We test four sets of regional features, each corresponding to one of the four regional feature categories, i.e., POI, structure, density, and community listed in Table 4. Fig. 15 shows the weight for each set. It is clear that the set of regional features in community is the relatively most important key factor set, and the regional features in structure are relatively less important key factors. The regional features in POI and density are not key factors.

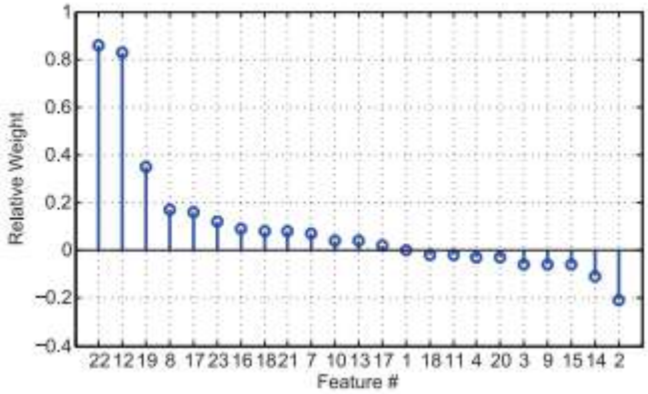


Fig. 16. The weights of the regional features in Table 4. From left to right: 1) house age, 2) # of neighbors, 3) road length / area size, 4) ratio of major / minor roads, 5) density of major roads, 6) house price, 7) density of leisures, 8) density of minor roads, 9) # of communities, 10) # of intersections, 11) average road length, 12) density of attractions, 13) density of major roads, 14) # of attractions, 15) # of minor roads, 16) area size, 17) # of leisures, 18) density of intersections, 19) # of hotels, 20) road length, 21) density of hotels, 22) density of restaurants, 23) # of restaurants.

8. CONCLUSION

This work fills the gap in the study on the impact of weather to traffic from few locations to all road networks throughout a city, more importantly, the regional features leading to the vulnerability of traffic in local areas to inclement weather are systematically revealed for the first time.

The empirical study in Shanghai demonstrates the effectiveness of the proposed system. The regional weather-traffic indices extracted have been validated to be surprisingly consistent with real world observations. Further regional key factor analysis yields interesting results.

For example, the regional house age has significant impact on the region's weather-traffic index. The achievement in this work will benefit government agent to understand the functional character of districts throughout a city, to improve traffic prediction and to learn the key factors in urban planning, etc.

The knowledge of key factors learned from one city is transferable to another city because modern cities often have road networks with similar quantitative density and other features. At last, the investigated problem has important practical value, but the research is still in its early stage. We are working on to incorporating more data sources to continuously improve the results.

REFERENCES

[1] Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni, "Dissecting regional weather-traffic sensitivity throughout a city," in Proc. 15th IEEE Int. Conf. Data Mining, 2015, pp. 739–744.

[2] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," ACM Trans. Intell. Syst. Technol., vol. 5, 2014, Art. no. 38.

[3] A. Saegusa and Y. Fujiwara, "A study on forecasting road surface conditions based on weather and road surface data," IEICE Trans. Inf. Syst., vol. 90-D, no. 2, pp. 509–516, 2007.

[4] M. A. Abdel-Aty and R. Pemmanaboina, "Calibrating a real-time traffic crash-prediction model using archived weather and its traffic data," IEEE Trans. Intell. Transp. Syst., vol. 7, no. 2, pp. 167–174, Jun. 2006.

[5] S. Dunne and B. Ghosh, "Weather adaptive traffic prediction using neurowavelet models," IEEE Trans. Intell. Transp. Syst., vol. 14, no. 1, pp. 370–379, Mar. 2013.

[6] M. J. Koetse and P. Rietveld, "The impact of climate change and weather on transport: An overview of empirical findings," Transp. Res. Part D: Transport Environment, vol. 14, no. 3, pp. 205–221, 2009.

[7] Talking about smart transportation from the rainstorm of Beijing, 2012. [Online]. Available: http://info.secu.hc360.com/2012/07/300822649026.shtml

[8] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-Air: When urban air quality inference meets big data," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 1436–1444.

[9] S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 919–928.

[10] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 186–194