Chapter 2
Early Warning of Human Crowds Based on Query Data from Baidu Maps: Analysis Based on Shanghai Stampede

Jingbo Zhou, Hongbin Pei, and Haishan Wu

Abstract Without sufficient preparation and on-site management, the mass-scale unexpected huge human crowd is a serious threat to public safety. A recent impressive tragedy is the 2014 Shanghai Stampede, where 36 people died and 49 were injured in celebration of the New Year's Eve on December 31st 2014 in the Shanghai Bund. Due to the innately stochastic and complicated individual movement, it is not easy to predict collective gatherings, which can potentially leads to crowd events. In this chapter, with leveraging the big data generated on Baidu Maps, we propose a novel approach to early warning such potential crowd disasters, which has profound public benefits. An insightful observation is that, with the prevalence and convenience of web map service, users usually search on the Baidu Maps to plan a routine, which reveals the users' future destinations. Therefore, aggregating users' query data on Baidu Maps can obtain priori and indication information for estimating future human population in a specific area ahead of time. Our careful analysis and deep investigation on the Baidu Maps' data on various events also demonstrate a strong correlation pattern between the number of map query and the number of positioning in an area. Based on such observation, we propose a decision method utilizing query data on Baidu Maps to invoke warnings for potential crowd events about 1~3 h in advance. Then we also construct a machine learning model with heterogeneous data (such as query data and positioning data) to quantitatively measure the risk of the potential crowd disasters. We evaluate the effectiveness of our methods on the data of Baidu Maps.

Keywords Emergency early warning • Crowd anomaly prediction • Map query • Stampede

---

J. Zhou • H. Wu (✉)
Baidu Research, Big Data Lab, Beijing, China
e-mail: hswu85@gmail.com

H. Pei
Baidu Research, Big Data Lab, Beijing, China

College of Computer Science and Technology, Jilin University, Changchun, China

Z. Shen, M. Li (eds.), *Big Data Support of Urban Planning and Management*,
Advances in Geographic Information Science, DOI 10.1007/978-3-319-51929-6_2

2.1 Introduction

The management of human crowd in public events is significantly important for public safety. The 2014 Shanghai Stampede, where 36 people died and 49 were injured in celebration of the New Year's Eve on December 31st 2014, was blamed for insufficient preparation and poor on-site management of local officials (Investigation Report 2014). Similar to other crowd disasters, one of major reasons of this tragedy is due to the wrong estimation of the human flow and human density by management officials (Investigation Report 2014). As the potential risk is increasing rapidly with the increment of mass-scale, unexpected huge crowd, inadequate emergency response preparedness may lead to a disaster. It is clear that the primary factor in assuring a safe environment for crowd is a feasible emergency action plan for potential crowd events, which can help preventing such terrible crowd disasters from happening.

Preventing crowd disasters is a challenging task which essentially depends on accurately predicting human movements. Although many existing works has modeled and predicted individuals' trajectory, most of them focus on human daily routine trajectory (Ashok 1996; Schneider et al. 2013; Zheng and Ni 2012). However, crowd anomaly is generally caused by infrequent collective activities (e.g., celebration, religious gatherings, sports tournaments). Participating these activities is a non-routine but stochastic behavior for most people. Thus, the movement of participators is much different from their daily routine and more like a random walk (Gonzalez et al. 2008).Therefore, such kind of movement can hardly be predicted readily, and there is no existing approach to handling it well by far, to the best of our knowledge.

Traditional method for crowd anomaly detection is based on video sensor and computer vision technology (Helbing et al. 2007; Gallup et al. 2012; Ali and Shah 2008). Whereas, three limitations of this kind of methods cannot be overcome: (1) they are too sensitive to noise of visual environment, e.g., rapid illumination changing may let them fail; (2) the areas with deployed video sensors are very limited in the citywide; and (3) the detection methods cannot provide significant long enough preparation time (e.g., hours level) for emergency management before crowd anomaly happens.

In this work, we propose an effective early warning method for crowd anomaly, as well as a machine learning model for crowd density prediction. Both of them are derived from a novel prospect by leveraging query data from Baidu Maps, the largest web mapping service application in China provided by Baidu.[1] Inspired by a widespread custom that people usually plan their trip on Baidu Maps before departure, we find a strong correlation pattern between the number of map query and the number of subsequent positioning users in an area. That is to say, map query behavior in some sense is a nice leading indicator and predictor for crowd dynamics. Therefore, we divide the early warning task into two parts: firstly, detecting map

[1] http://map.baidu.com/

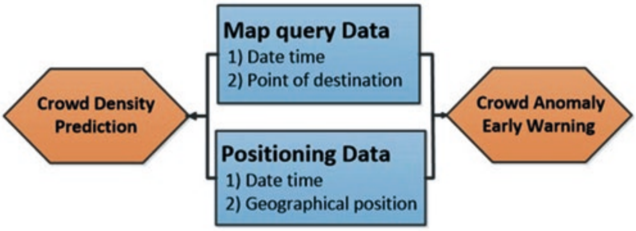


Fig 2.1 The framework of early warning for crowd anomaly based on map query data and positioning data

query dynamics of an area and sending early warning when there is a map query number anomaly, and then quantitatively predicting crowd density from historical map query data and positioning data of this area. The framework is illustrated in Fig. 2.1. Our approach can reliably provide early warning for crowd anomaly for 1–3 h which would facilitate an early intervention for administrative agency to prevent a crowd disaster from happening.

The main contributions of this work are listed as follows:

We provide an inspiring perspective to predict the crowd anomaly according to mass map query behavior.

We propose a simple but effective early warning method for crowd anomaly based on map query data and positioning data. Case studies and experiments on real data demonstrate that the proposed method can achieve an early warning for anomaly 1–3 h ahead.

We develop a machine learning model for crowd density prediction based on historical map query data and positioning data. Experiments on real data show that the model can accurately predict crowd density in a specific area one hour ahead.

The rest of this chapter is organized as follows. First, we give a detailed analysis of the 2014 Shanghai Stampede based on positioning data and map query data in Sect. 2.2. Then the details of the proposed model and the experimental results are described in Sect. 2.3. Finally, we conclude this work in Sect. 2.4.

2.2 Case Study of the Shanghai Stampede

2.2.1 Observations from the Data of the Shanghai Stampede

In this section, we present several important observations from the human mobility data about the 2014 Shanghai Stampede obtained from Baidu Maps, which can provide some insights for human flow and crowd anomaly prediction. The Shanghai Stampede happened at the Bund in Shanghai on December 31st 2014. Hereafter, we use the “map query number” during a time interval (typically it is one hour) to denote the total number of queries whose destination is located in a specific area

Fig. 2.2 Human population density between 23:00 and 24:00 on December 31st 2014

(like the Shanghai Bund) on Baidu Maps. Meanwhile, we denote the number of users by enabling the positioning function of Baidu Maps during a time interval in a specific area as the “positioning number”. The position number can be treated as an approximation of the human population density which is the number of persons per square within an area during a fixed time interval.

Observation 1: Human Population Density of the Stampede Area Was Higher than Others
The high human population density is a necessary condition for the happening of stampede. Figure 2.2 illustrates the human density heat map between 23:00 and 24:00 on December 31st 2014 based on the positioning number. It is obvious that the Bund, which is the disaster area of 2014 Shanghai stampede, had higher human density than the rest parts of the city.

Observation 2: Human Population Density of the Disaster Area Was Higher than the Ones at Other Times
The human population density when the disaster happened was also higher than the ones at other times. Figure 2.3 shows the positioning number on Baidu Maps on the Bund over one week. During the disaster time, the positioning number on Baidu Maps was about 9 times of peak values at other times. Furthermore, Fig. 2.4 exhibits that the user positioning number on the Shanghai Bund of the New Year’s Eve of 2014 was also higher than other festivals/holidays (in Fig. 2.4, the festivals/holidays are August 23rd (a common weekend of 2014), the Mid-Autumn Eve (September 7th the evening before a traditional Chinese festival), the National Day on (October 1st), and the New Year’s Eve (December 31st)).

Observation 3: The Human Flow Directions of the Disaster Area Were More Chaotic
We also observe that the human flow directions in the disaster time were more chaotic than the ones at other times. Figure 2.5 illustrates the human flow directions in different festivals/holidays from 22:00 to 24:00 in the disaster area. Compared with other subfigures, Fig. 2.5(d) indicates that the disaster area had more chaos during the New Year’s Eve of 2014, which implies a potential accident.

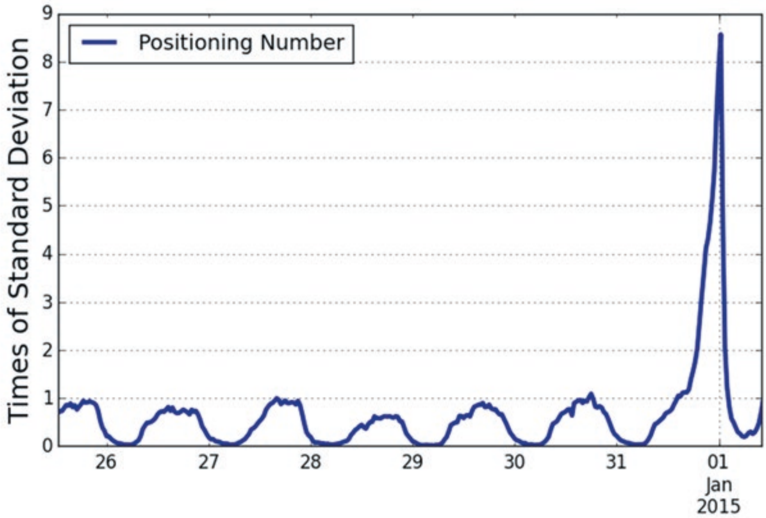


Fig. 2.3 The user positioning number in Baidu Maps on the Bund (the disaster area of 2014 Shanghai Stampede) over one week

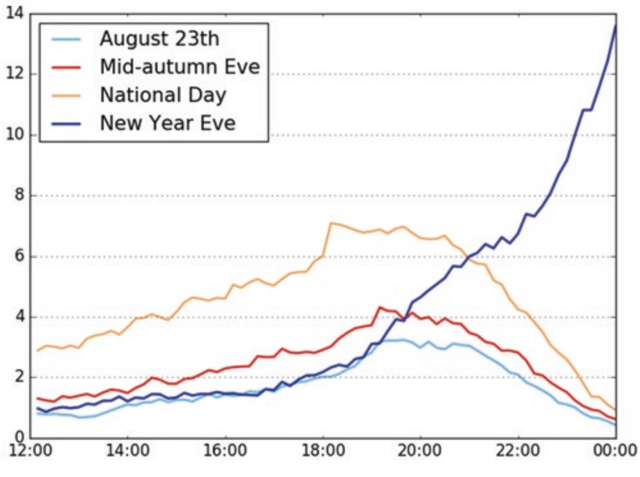


Fig. 2.4 The positioning number in Baidu Maps on the Bund (the disaster area of 2014 Shanghai Stampede) in different festivals/holidays of 2014, which are August 23rd (a common weekend of 2014), the Mid-Autumn Eve (7 September, the evening before a traditional Chinese festival), the National Day (October 1st 2014), and the New Year's Eve of 2014. All the numbers are divided by the standard deviation of the position number of Mid-Autumn Eve

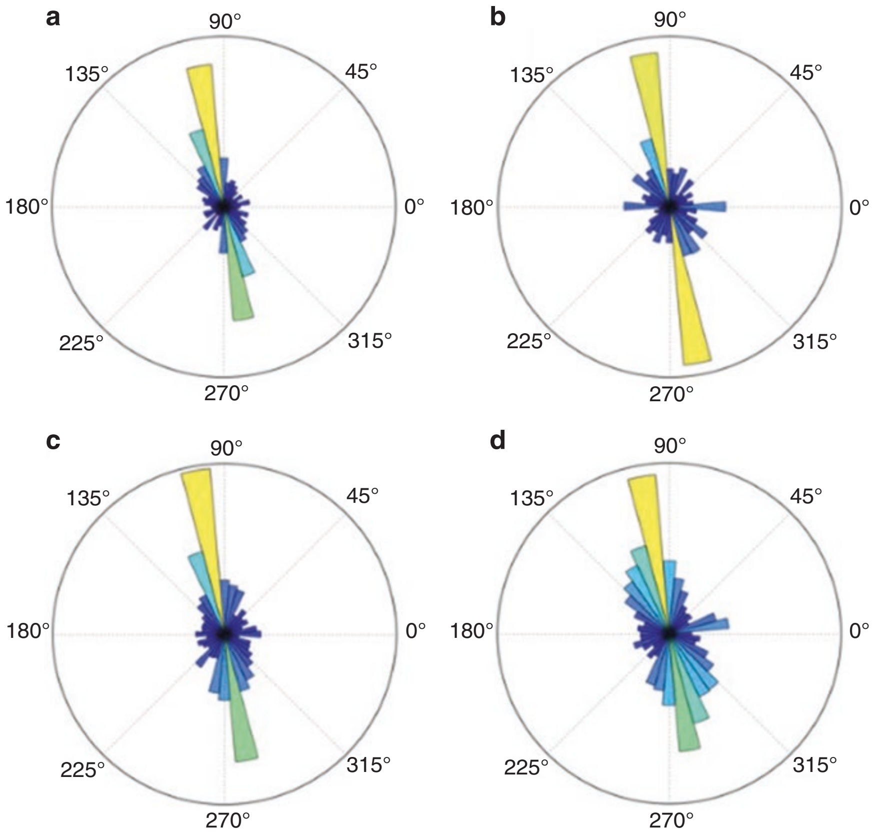


Fig. 2.5 Human flow direction distribution in Chenyi Square (the specific disaster area of 2014 Shanghai Stampede) from 22:00 to 24:00 in: (**a**), common weekend (August 23rd 2014); (**b**), the eve of the Mid-Autumn Festival (September 7th 2014); (**c**), the China's National Day (October 1st 2014); and (**d**), New Year's Eve of 2014

2.2.2 How Baidu's Data Can Help Prevent Crowd Disasters

Our objective is to utilize the big data, especially the query data, generated by Baidu Maps' users to avoid such crowd disasters. If without prior information, the individual movement appears as a random walk (Gonzalez et al. 2008), leading to the unavoidable impediment to predict unknown human crowd event. However, nowadays, before a user leaves for a destination, he or she usually uses a web mapping service application to plan a routine to the destination. These queries on the web map reveal some information about users' future behavior in a few of hours. Taking more than 70% market share overall in China, the Baidu Maps has innate advantage to tackle this problem. Therefore, collecting and aggregating the massive query data on Baidu Maps provides an intelligent approach to predicting crowd anomaly as well as avoiding such crowd disasters.

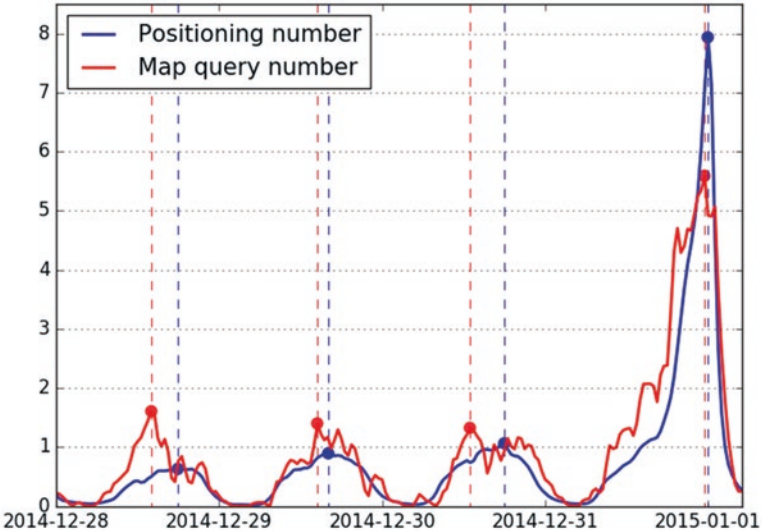


Fig. 2.6 The time series of map query number and the positioning number on Baidu Maps from December 28th 2014 to December 31st 2014 on the Shanghai Bund (both data are normalized by dividing their standard deviation). An interesting observation is that the peak values of the map query number usually appear several hours before the peak value of the positioning number in each day

After analyzing the data on the Shanghai Stampede and other crowd events (like game events and concerts) in different places, we find that there usually are an abnormally large number of queries on the Baidu Maps about the crowd event places emerging 0.5–2 h before the abnormal crowd event being perceptible to human beings.

Figure 2.6 illustrates the time series of positioning number and the one of the map query number of the Baidu Maps in the last a few of days of 2014. This figure exposures the high correlation between the positioning number and the map query number. With a more careful inspection, we can also observe that the peak value of the map query number appears, generally, several hours before the peak value of the positioning number in each day.

We expound more investigations to justify the map query number as a leading indicator of human flow from the practical and theoretical perspectives, which are introduced as follows:

First, as shown in Fig. 2.7, we exhibit the distribution of the time lags (in hours) between the peak value of the map query number and the positioning number on the Shanghai Bund in each day over one year. The negative lag means the peak value of the map query number appears ahead of the one of the positioning number. As we can see from Fig. 2.7, about 80% of days in 2014, the peak value of the map query number appears one hour before the peak value of the positioning number, and with

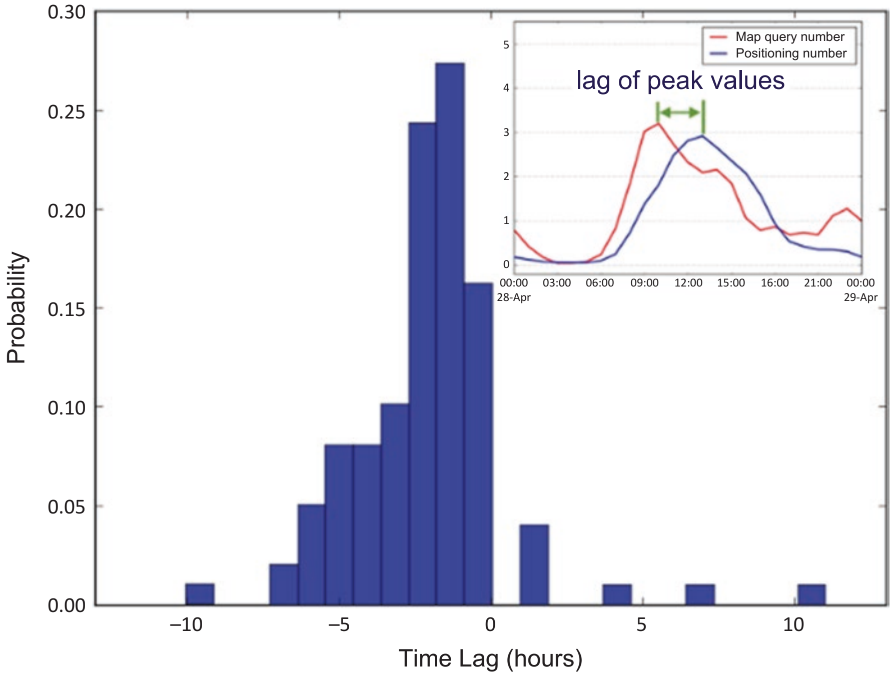


Fig. 2.7 The distribution of the lags (in hours) of the peak values between the map query number and the positioning number within the area of the Shanghai Bund in 2014. The negative lag means the peak of the map query number appears ahead of the positioning number

more than 50% of days in 2014, the peak value of the map query number appears two hours before the one of the positioning number.

Second, Fig. 2.8 theoretically demonstrates that the map query number can help predict the future human flow with the mutual information analysis. As shown in Fig. 2.8, knowing the map query number with one hour ahead can reduce the uncertainty (aka the entropy) of the positioning number, which can empower our ability to predict the future human flow in an area.

2.2.3 The General Usability of Baidu Maps' Data

We also explore the relations between the map query number and the positioning number on Baidu Maps in many different places to demonstrate the general usability of the map query data of Baidu Maps for predicting crowd anomaly. In this section, we investigate three types of such points of interest (POIs) in Beijing, which are: public event place, landmark, and transportation node. The selected representative places of them are the Beijing Workers' Stadium, the Forbidden City, and the Beijing West Railway Station, respectively.

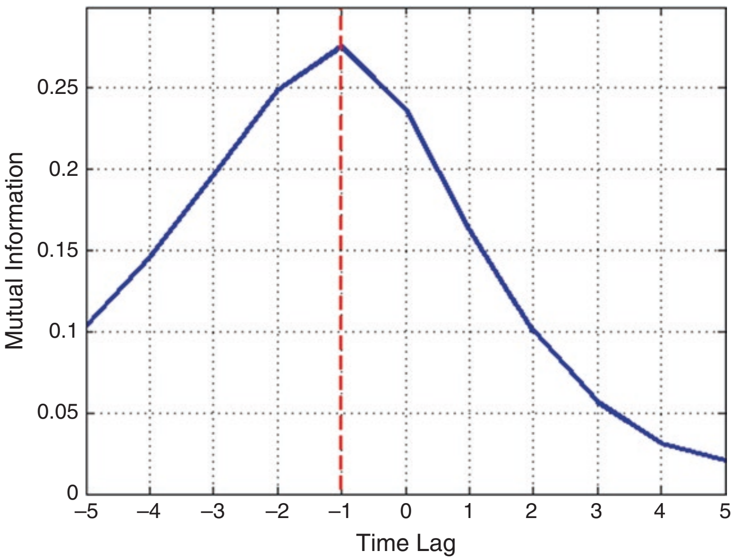
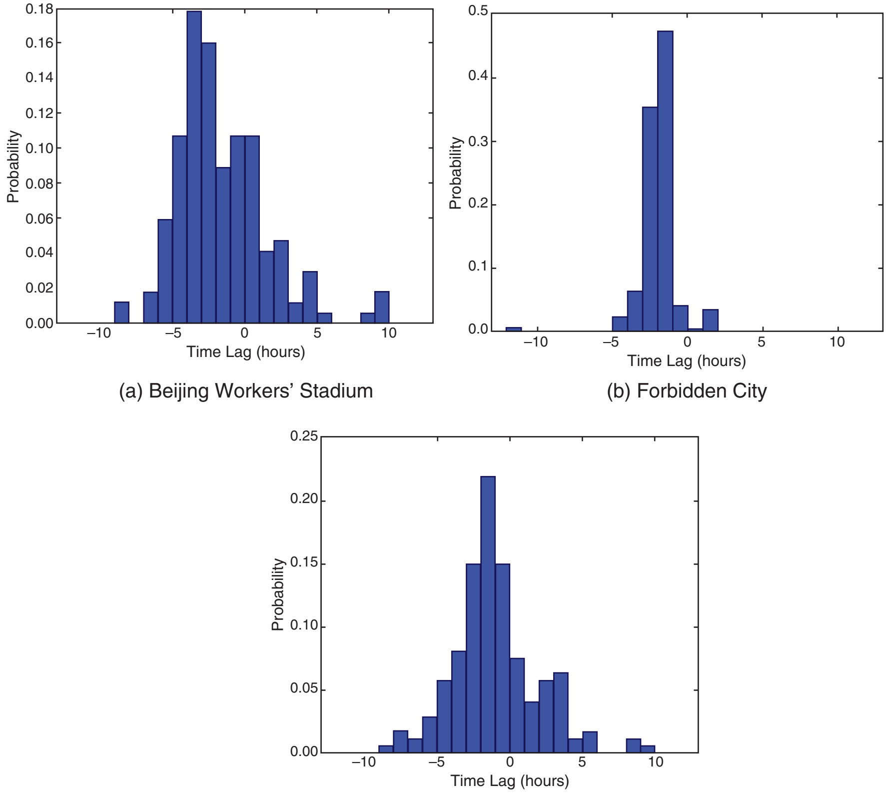


Fig. 2.8 The mutual information between the time series of the map query number and the positioning number within the area of the Shanghai Bund with different lags (in hours) in 2014

The statistical relation analysis, which is the same with the ones in Figs. 2.7 and 2.8 of Sect. 2.2.2, is exhibited in Figs. 2.9 and 2.10. All the analysis demonstrates that the map query number can help predict the crowd anomaly in an area. It is also worthwhile to note that different places may also have different characters for such relations. For example, in the Forbidden City, the peaks of map query number concentrate on two and three hours before the peaks of the positioning numbers. A more interesting observation of the Forbidden City is shown in Fig. 2.11. In the Forbidden City, for each day, there are two peaks of the time series of the map query number before the peak of the positioning number emerges. This reflects the complex dynamics of human behaviors and diverse environment conditions of different places.

2.3 Preventing Crowd Disaster with Baidu Maps' Data

In this section, we detail how to utilize the Baidu Maps' data to prevent the potential crowd disasters. We first present a qualitative decision method for crowd anomaly prediction solely based on the map query data on Baidu Maps. Then we construct a machine learning model with heterogeneous data to quantitatively measure the risk of the potential crowd disasters.



(a) Beijing Workers' Stadium

(b) Forbidden City

(c) Beijing West Railway Station

Fig. 2.9 The distribution of the lags of the peak values between the map query number and the positioning number within the area of different POIs in 2014. The negative lag means the peak of the map query number appears ahead of the one of positioning number: (**a**) Beijing Workers' Stadium, (**b**) Forbidden City, (**c**) Beijing West Railway Station

2.3.1 A Decision Method for Crowd Anomaly Prediction with Map Query Data

The general idea of the decision method for crowd anomaly prediction is that, if the number of map query per hour of a specific area is larger than a warning line (a certain threshold), we will invoke an anomaly warning about the possible crowd event in the following several hours. This decision method is based on relation analysis between the map query number and the positioning number discussed in Sect. 2.2, which indicates that a large number of map query number subsequently imply a potential large number of human population in a few of hours in a specific area.

The warning line for the map query number is statistically computed with the peak values of the map query number per hour in every day under the log-normal probability distribution model. Formally, let us denote the time series of map query

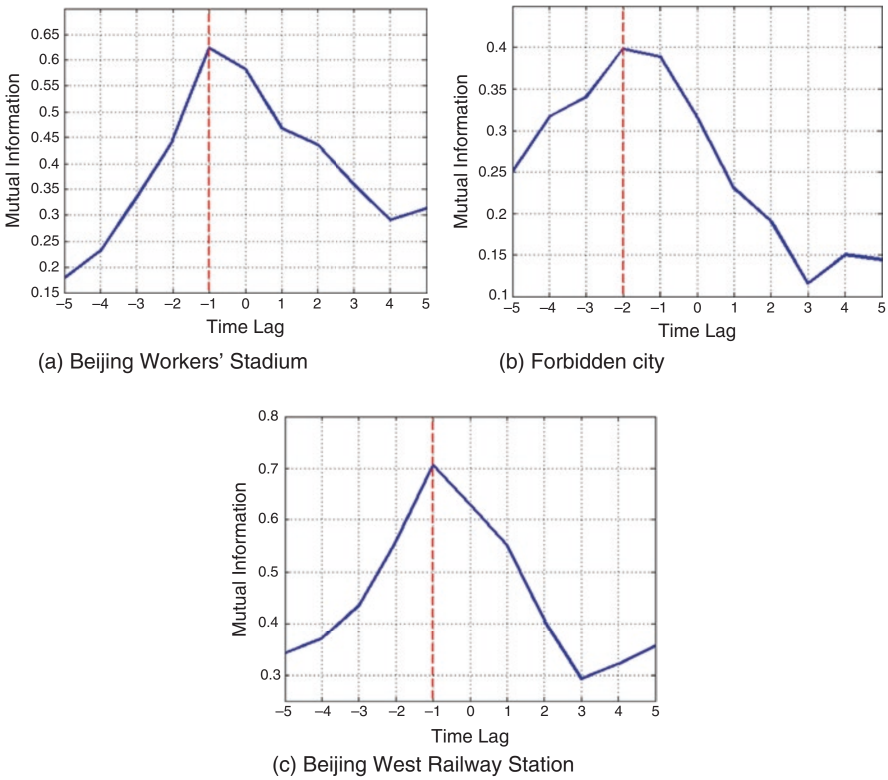


Fig. 2.10 The mutual information between the 1-year time series of the map query number and the positioning number within the area of different POIs in 2014: (**a**) Beijing Workers' Stadium, (**b**) Forbidden City, (**c**) Beijing West Railway Station

number per hour on the Baidu Maps by $M = m(t)$, $t \in \pm \mathrm{N}_0$. For a given date d, we first calculate the peak value of the map query number as pm(d):

$$pm(d) = \max_{t=24d}^{24(d+1)} m(t)$$

In our method, we consider pm(d) as random variable and assume it follows log-normal distribution, i.e.:

$$\varphi(d) = \log(\mathrm{pm}(d)) \sim N(\mu_{\mathrm{pm}}, \sigma_{\mathrm{pm}}^2) \tag{2.1}$$

We use one year time series data of map query number to obtain an unbiased estimation of μ_{pm} and σ_{pm} which are sample mean $\hat{\mu}_{\mathrm{pm}}$ and sample variance $\hat{\sigma}_{\mathrm{pm}}^2$:

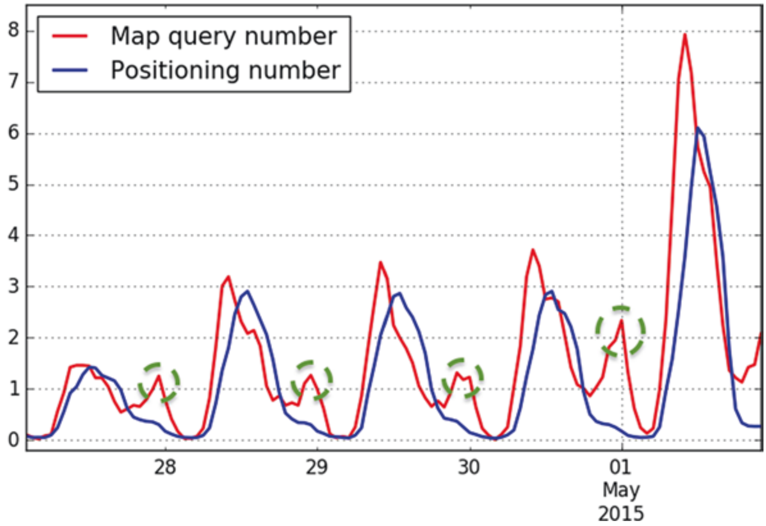


Fig. 2.11 The map query number and the positioning number of Baidu Maps' users from December 28th 2014 to December 31st 2014 in the Forbidden City (both data are normalized by dividing their standard deviation). For each day, the time series of the map query number has two peaks before the peak of positioning number

$$\hat{\mu}_{\mathrm{pm}} = \frac{1}{365} \sum_{d=0}^{364} \log\left(\mathrm{pm}(d)\right) \tag{2.2}$$

$$\hat{\sigma}_{\mathrm{pm}}^{2} = \frac{1}{365-1} \sum_{d=0}^{364} \left(\log\left(\mathrm{pm}(d)\right) - \hat{\mu}_{\mathrm{pm}}\right)^{2} \tag{2.3}$$

Accordingly, the warning line of the map query number is determined as:

$$\omega_{m} = \hat{\mu}_{\mathrm{pm}} + \alpha \hat{\sigma}_{\mathrm{pm}} \tag{2.4}$$

where α is a model parameter and $\alpha > 0$. The effect of α is evaluated in Sect. 2.3.1.2.

Similarly, we can also obtain a warning line of the positioning number. We define the time series of position number as $Q = \{q(t)\}$, and the peak value of positioning number for date d is pq(d):

$$\mathrm{pq}\left(d\right) = \max_{t=24d}^{24(d+1)} q(t)$$

We also assume that pq(d) follows log-normal distribution, i.e.:

$$\phi(d) = \log\left(\mathrm{pq}(d)\right) \sim N\left(\mu_{\mathrm{pq}}, \sigma_{\mathrm{pq}}^2\right) \tag{2.5}$$

and we have the sample estimation of $\hat{\mu}_{\mathrm{pq}}$ and $\hat{\sigma}_{\mathrm{pq}}^2$:

$$\hat{\mu}_{\mathrm{pq}} = \frac{1}{365}\sum_{d=0}^{364} \log\left(\mathrm{pq}(d)\right) \tag{2.6}$$

$$\hat{\sigma}_{\mathrm{pq}}^2 = \frac{1}{365-1}\sum_{d=0}^{364} \left(\log\left(\mathrm{pq}(d)\right) - \hat{\mu}_{\mathrm{pq}}\right)^2 \tag{2.7}$$

In this chapter, the warning line of positioning number is fixed as $\omega_q = \hat{\mu}_{\mathrm{pq}} + 3\hat{\sigma}_{\mathrm{pq}}$. Finally, our decision method can be formally described by following lemma:

Lemmma 2.1 If $m(t) \geq \omega_m$, then we have $q(t + \Delta) \geq \varpi_q$ and $1 \leq \Delta \leq T$ with high probability, where T is a limited time period.

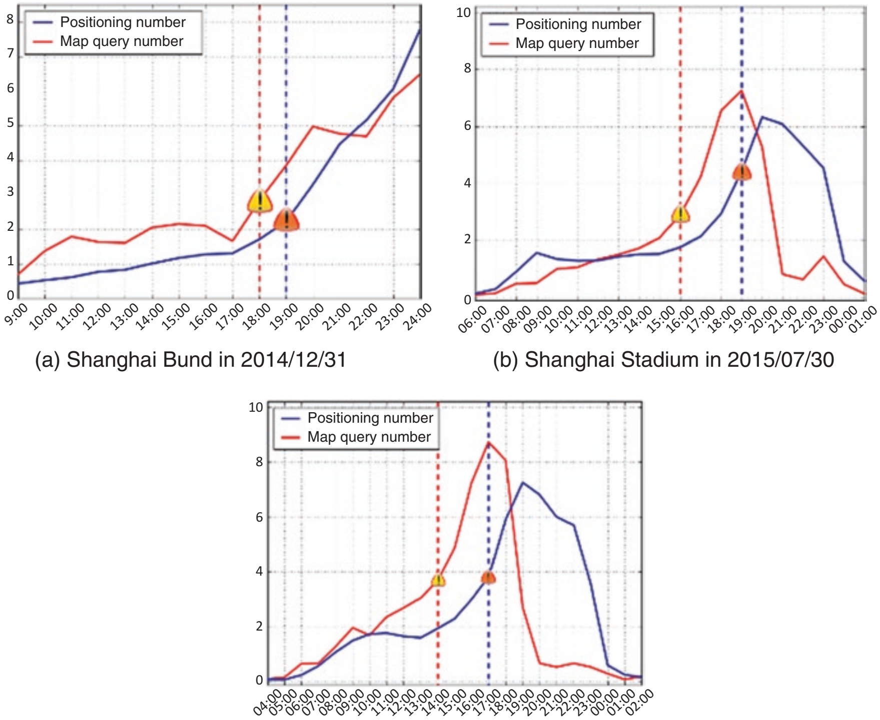
In our following demonstrations and evaluations, we set T = 3 (hours).

2.3.1.1 Demonstration of the Decision Method

We demonstrate several examples for crowd event prediction based on Lemma 2.1 in different POIs of several China cities. Figure 2.12 illustrates such decision method on the data about the public crowd anomaly prediction of Shanghai and Shenzhen (which can be compared with Fig. 2.13). As we can see from Fig. 2.12, a warning about the abnormal crowd event can be invoked 1–3 h before the human population surpassing a warning line. Figure 2.14 demonstrates similar results for the abnormal crowd event prediction in Beijing. To sum up, with monitoring the map query number for each hour, it is possible to give a very early warning for the abnormal crowd event to avoid crowd disasters (Fig. 2.15).

2.3.1.2 Performance Evaluation for the Decision Method

We also employ a performance evaluation for the decision method. In the six POIs of Figs. 2.12 and 2.14, we first select all the time points where $q(t) > \omega_q$ and consider these points as the ground truth of crowd events. Then we use our decision method introduced in Lemma 2.1 to try to predict the possible crowd events. In other words, for a monitored POI, if the map query number is larger than a predefined warning line, we will invoke a warning that, in the following 1–3 h, there will be a crowd event. We exhibit the precision, recall, and F1-score of six POIs in Fig. 2.16. As we can see from the Fig. 2.16, there is a trade-off between the precision and recall, which is controlled by the times of the standard deviations (i.e., α in Eqn. 2.4). We also aggregate the precision, recall, and F1-score of the six POIs together and illustrate the average measures in Fig. 2.17.



(a) Shanghai Bund in 2014/12/31

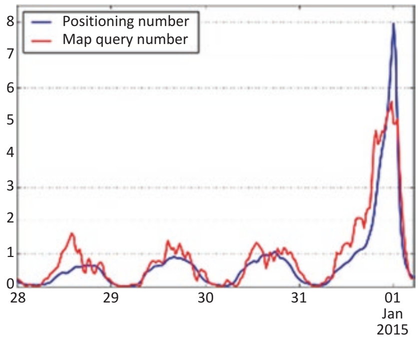
(b) Shanghai Stadium in 2015/07/30

(c) Shanghai Bay Sports Center in 2015/04/25

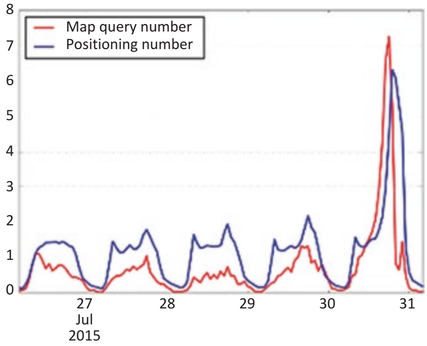
Fig. 2.12 An illustration of the decision method for abnormal crowd event detection with map query data on Baidu Maps in Shanghai and Shenzhen. If the map query number is larger than a warning line, we will invoke a warning for crowd event in the following several hours (see Lemma 2.1). From the figure we can see that the event warning can be sent 1–3 h before the human population surpassing the warning line. The corresponding events are (**a**) 2014 Shanghai Stampede (Shanghai Bund in 2014/12/31), (**b**) the International Champions Cup Real Madrid vs AC Milan (Shanghai Stadium in 2015/07/30) (http://www.goal.com/en-za/match/real-madrid-vs-milan/2032786/report), and (**c**) TVXQ Special Live Tour T1story in Shenzhen (Shenzhen Bay Sports Center in 2015/04/25) (http://www.springcocoon.com/html/CN/ssyc/hdgl/201504/07847.html)

2.3.2 Machine Learning Model for Risk Control of the Potential Crowd Disasters

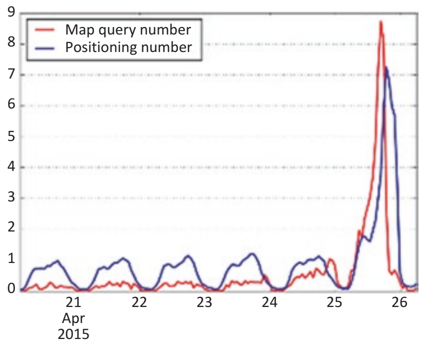
We also develop a machine learning model to utilize the map query data, historical positioning data, and other information available in the Baidu Maps to predict the future human density in an area. Such predicted density can be taken as a quantitative measure to assess the potential risk of a crowd disaster. The intuition is that the probability of a crowd disaster is directly related with the human density and the human density is a necessary (though not sufficient) condition for a crowd disaster.



(a) Shanghai Bund in 2014/12/31



(b) Shanghai Stadium in 2015/07/30



(c) Shenzhen Bay Sports Center in 2015/04/25

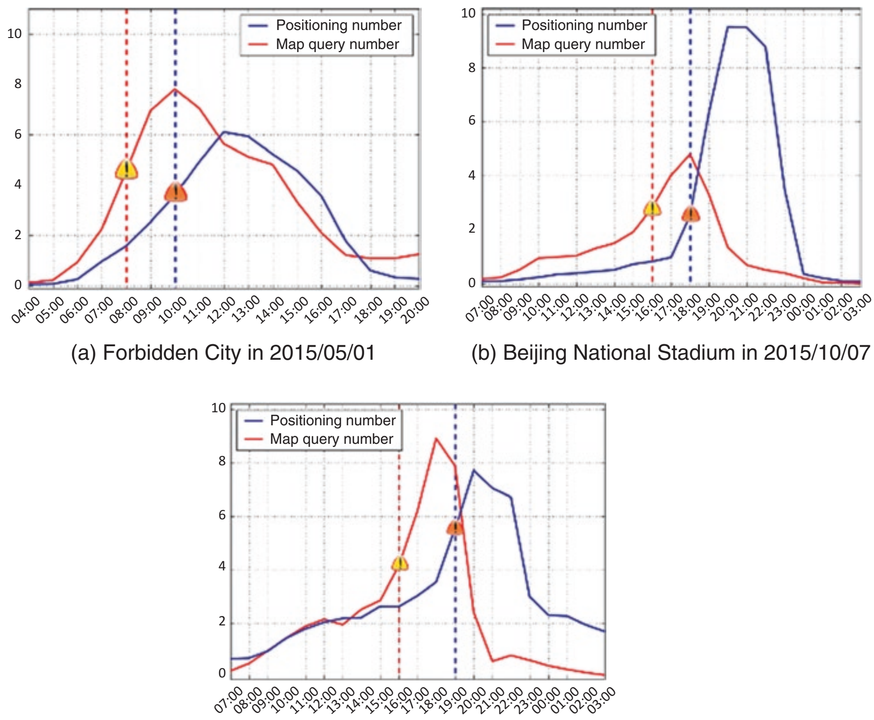
Fig. 2.13 The map query number and positioning number on the Baidu Maps corresponding to the events of Shanghai and Shenzhen shown in Fig. 2.12: (**a**) Shanghai Bund on in 2014/12/31, (**b**) Shanghai Stadium on 2015/07/30, (**c**) Shenzhen Bay Sports Center in 2015/04/25

2.3.2.1 Problem Definition

Let us denote $M_{t_c} = m(t)$ and $Q_{t_c} = q(t), t \in [0,,\ldots,,t_c]$ as time series of map query number and time series of positioning number before time t_c in a specific area, respectively. We set time granularity to one hour with a fine resolution and statistical significance. Let $f\left(M_{t_c}, Q_{t_c}\right)$ denote the prediction model with two time series of input; then the positioning number at next time $q(t_c + 1)$ can be estimated as:

$$q\left(t_c + 1\right) \approx \hat{q}\left(t_c + 1\right) = f\left(M_{t_c}, Q_{t_c}\right) \tag{2.8}$$

where $q\left(t_c + 1\right) \approx \hat{q}\left(t_c + 1\right)$ means approximate. Thus, the problem what we are interested in is: can we train a prediction model f, from the historical map query and positioning data, to accurately predict the positioning number in the next time $q(t_c+1)$?



(a) Forbidden City in 2015/05/01

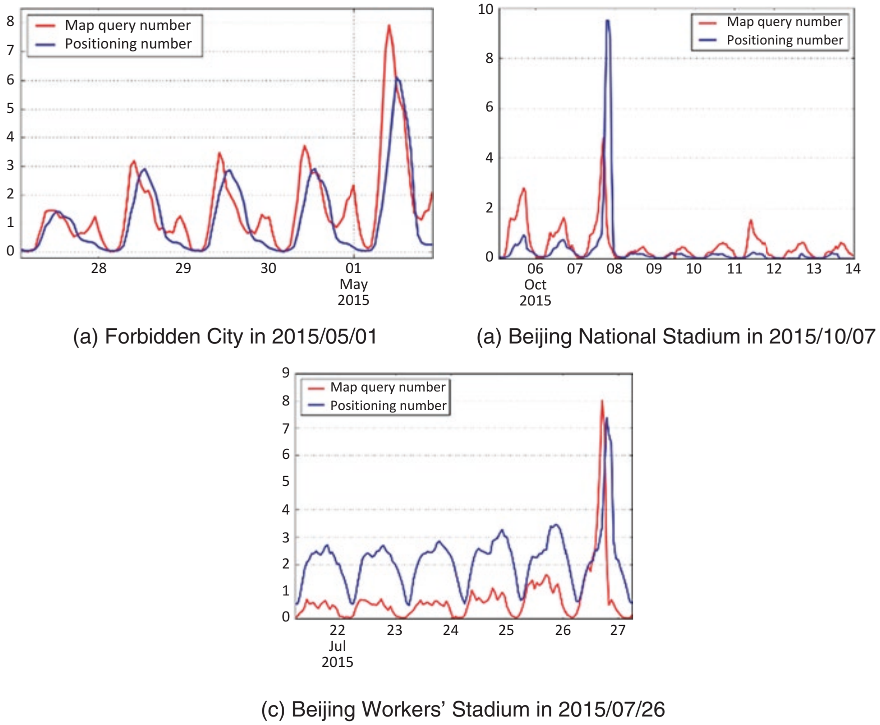
(b) Beijing National Stadium in 2015/10/07

(c) Beijing Workers' Stadium in 2015/07/26

Fig. 2.14 An illustration of the decision method for abnormal crowd event detection with map query data on Baidu Maps in Beijing (similar with Fig. 2.12). If the map query number is larger than a warning line, we will invoke a warning for crowd event in the following several hours (see Lemma 2.1). From the figure, we can see that the event warning can be sent 1–3 h before the positioning number surpassing the warning line. The corresponding events are (**a**) International Workers' Day (Forbidden City in 2015/05/01), (**b**) Finals of the Voice of China (Beijing National Stadium in 2015/05/01, http://www.n-s.cn/ssyc/1/), and (**c**) the Hunting Party Chinese Tour (Beijing Workers' Stadium in 2015/07/26, http://lplive.net/shows/db/2015/20150726)

2.3.2.2 Model and Feature Selection

We model the positioning number prediction task as a regression problem, where the positioning number in next hour $q(t_c+1)$ is the target value. A gradient boosting decision tree (GBDT) model, which is an effective and widely adopted ensemble machine learning technique, is employed for the task. In the GBDT model, we extract 47 features from two time series M_{tc} and Q_{tc} for the prediction task; the details of the features are described in Table 2.1.



(a) Forbidden City in 2015/05/01

(a) Beijing National Stadium in 2015/10/07

(c) Beijing Workers' Stadium in 2015/07/26

Fig. 2.15 The map query number and positioning number on the Baidu Maps corresponding to the events of Beijing shown in Fig. 2.14: (**a**) Forbidden City in 2015/05/01, (**b**) Beijing National Stadium in 2015/10/07, (**c**) Beijing Workers' Stadium in 2015/07/26

2.3.2.3 Experiments

For the aforementioned six different POIs (Shanghai Bund, Shanghai Stadium, Shenzhen Bay Sports Center, Forbidden City, Beijing National Stadium, Beijing Workers' Stadium), we train six distinctive positioning number prediction models for them based on the data in 60 days before the event in each POI. As we can see from Fig. 2.18, the predicted positioning number of our model is very near to the real number. That is to say, the model provides a reliable one hour ahead prediction in the six POIs. With high quality positioning data and map query data of Baidu, we believe the positioning number prediction model can be extended to more other POIs.

Furthermore, we evaluated the importance of all features in prediction model. The top 10 important features are demonstrated in Fig. 2.19 (feature description in Table 2.1). We calculate the Gini importance score (the higher the score, the more important the feature) for each feature (L. Breiman 2001). As the figure shows, "PN1 (positioning number 1 hour ago)" and "MQ1 (map query number 1 hour ago)" are the two most important features in our model, which is consistent with our

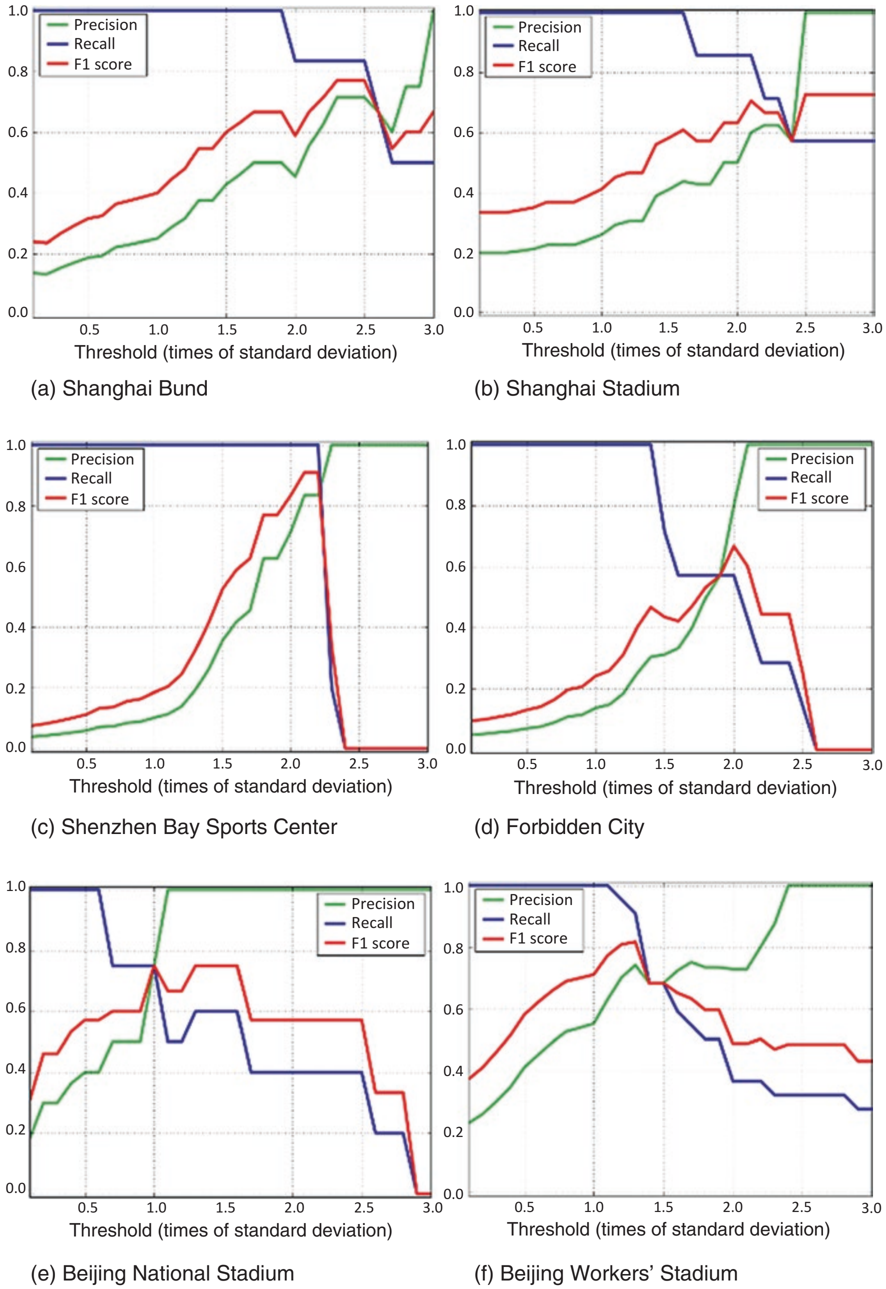


Fig. 2.16 The precision, recall, and F1-score for abnormal crowd event prediction with map query data in different places with different times (i.e., α) of standard deviations: (**a**) Shanghai Bund, (**b**) Shanghai Stadium, (**c**) Shenzhen Bay Sports Center, (**d**) Forbidden City, (**e**) Beijing National Stadium, (**f**) Beijing Workers' Stadium

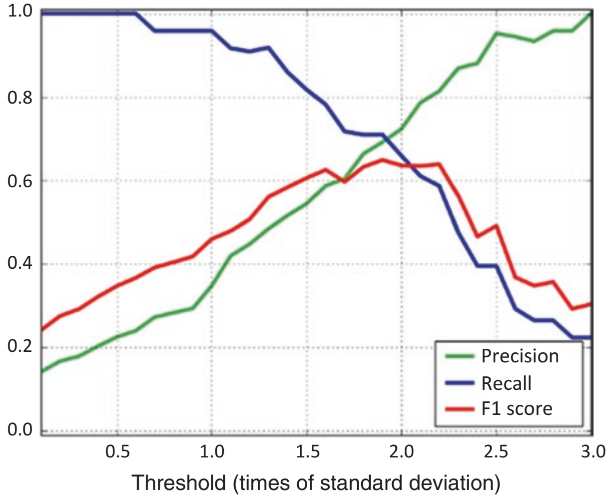
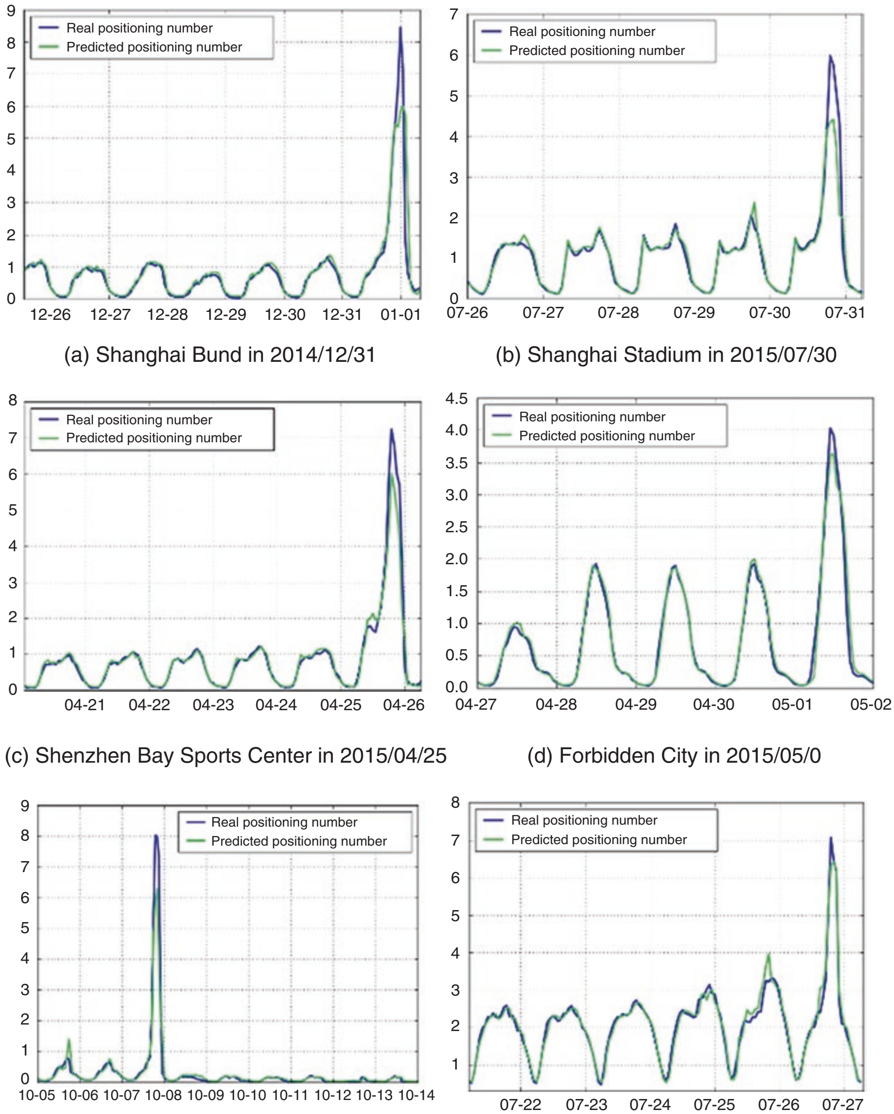


Fig. 2.17 The precision, recall, and F1-score for crowd anomaly prediction with map query data averaged in six POIs (which are the Shanghai Bund, Shanghai Stadium, Shenzhen Bay Sports Center, Forbidden City, Beijing National Stadium, and Beijing Workers' Stadium)

Table 2.1 Feature Instruction

ID	Name	Feature description
1	PN1	Positioning number 1 hago
2	PN2	Positioning number 2 h ago
….	..	…
4	PN4	Positioning number 4 h ago
5	PNS1	Positioning number at same hour 1 day ago
…	…	…
11	PNS7	Positioning number at same hour 7 day ago
12	MQ1	Map query number 1 h ago
13	MQ2	Map query number 2 h ago
14	MQY	Map query number in 20:00–24:00 yesterday
15	CT1	Current time is 1 o'clock (bool value)
…	…	…
38	CT24	Current time is 24 o'clock (bool value)
39	TW1	Today is Monday (bool value)
…	…	…
45	TW7	Today is Sunday (bool value)
46	WD	Today is weekend (bool value)
47	HD	Today is holiday (bool value)



(a) Shanghai Bund in 2014/12/31 (b) Shanghai Stadium in 2015/07/30

(c) Shenzhen Bay Sports Center in 2015/04/25 (d) Forbidden City in 2015/05/0

(e) Beijing National Stadium in 2015/10/07 (f) Beijing Workers' Stadium in 2015/07/26

Fig. 2.18 The predicted positioning number and real positioning number on the Baidu Maps corresponding to the aforementioned events: (**a**) Shanghai Bund in 2014/12/31, (**b**) Shanghai Stadium in 2015/07/30, (**c**) Shenzhen Bay Sports Center in 2015/04/25, (**d**) Forbidden City in 2015/05/01, (**e**) Beijing National Stadium in 2015/10/07, (**f**) Beijing Workers' Stadium in 2015/07/26

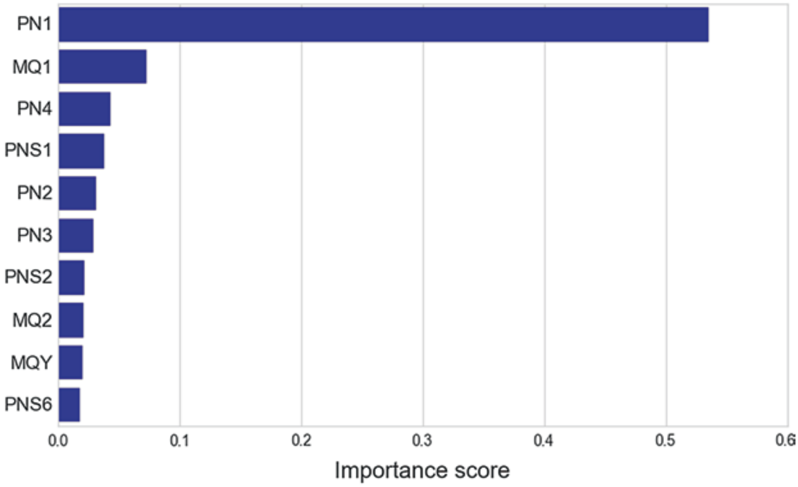


Fig. 2.19 The feature importance rankings in prediction model

intuition. It is worth noting that "MQY (map query number in 20:00–24:00 yesterday)" is also an important feature as many people prefer to plan their routine in the night before.

From the importance ranking in Fig. 2.19, we can see features from map query data are a vital part for positioning number prediction. In order to test the essentiality of map query data for prediction, we further design a comparative experiment, where two GBDT prediction models are developed: one with features from map query data and the other without. Then, we can directly capture the importance of map query data by quantity measuring the prediction error of two models. We use mean absolute error (MAE) as evaluation metrics to measure how close predictions are to the real positioning numbers. As the experiment result shown in Fig. 2.20, the performance of the model with the features is superior than the model without the features. To sum up, map query data exactly play a vital role in positioning number prediction.

2.4 Conclusion

The management of unexpected huge human crowd events is a challenging but serious problem for public health, whereas poor management of such events may lead to unfortunate stampedes. With the help of the big data generated on the Baidu Maps, we propose a solution for improving the effectiveness of the crowd management. The novelty of the solution lies in an objective fact that many people will search on Baidu Maps to plan their itineraries. This prior information indicated by anonymous users of Baidu Maps provides promising and compelling advantages for

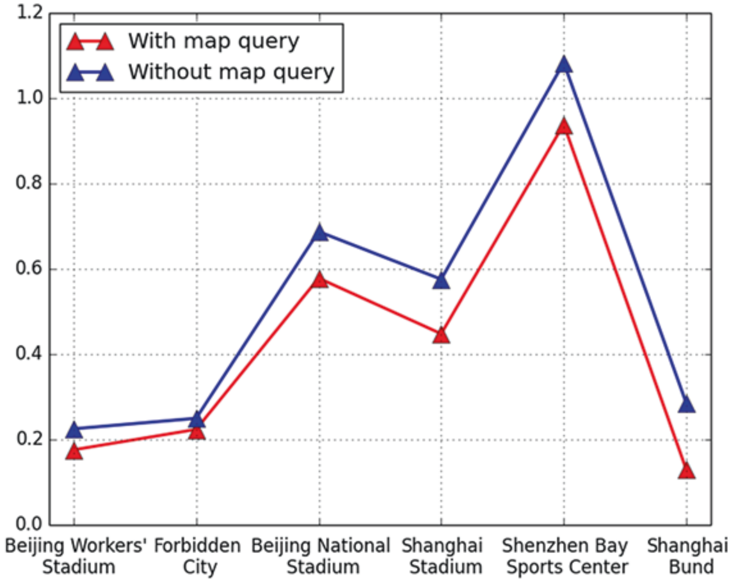


Fig. 2.20 The Mean Absolute Error (MAE) of the prediction models. *Red* line is the result of model with features in Table 2.1, and *blue* line is the one with features in Table 2.1 but without map query related features; y denotes the MAE score

Fig. 2.21 A snapshot of the early warning system for human crowds

us to invoke early warning for possible crowd events. In this chapter, we deploy a deep study for the crowd anomaly prediction with resorting to the help of the Baidu Maps' data. Our solution includes a qualitative decision method to perceive the crowd anomaly as well as a quantitative prediction method to assess the risk of the crowd event. We believe the successful deployment of our method can bring many benefits to our society.

It is worth mentioning that we have developed an early warning system for human crowds based on the aforementioned decision method and prediction model. A completed demonstration video of the system can be visited,[2] and a snapshot is shown in Fig. 2.21.

References

Investigation report of 2014 shanghai stampede event. http://www.shjcw.gov.cn/2015jjw/n2230/n2237/ u1ai51007.html

Ali, S., & Shah, M. (2008). Floor fields for tracking in high density crowd scenes. In *Computer Vision–ECCV 2008* (pp. 1–14). Heidelberg: Springer.

Ashok, K. (1996). *Estimation and prediction of time-dependent origin-destination flows* (PhD thesis). Massachusetts Institute of Technology.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Gallup, A. C., Hale, J. J., Sumpter, D. J., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences, 109*(19), 7245–7250.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Helbing, D., Johansson, A., & Al-Abideen, H. Z. (2007). Dynamics of crowd disasters: An empirical study. *Physical Review E, 75*(4), 046109.

Schneider, C. M., Belik, V., Couronn'e, T., Smoreda, Z., & Gonza'lez, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society Interface, 10*(84), 20130246.

Zheng, J., & Ni, L. M. (2012). An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 153–162). ACM.

[2] http://www.iqiyi.com/w_19rrpwo1mt.html#vfrm=2-3-0-1