Article

Using Adverse Weather Data in Social Media to Assist with City-Level Traffic Situation Awareness and Alerting

Hao Lu [1,2,†], **Yifan Zhu [1,†]**, **Kaize Shi [1]**, **Yisheng Lv [2]**, **Pengfei Shi [1]** and **Zhendong Niu [1,3,*]**

[1] School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China; hao.lu@ia.ac.cn (H.L.); zhuyifan@bit.edu.cn (Y.Z.); kzshi@bit.edu.cn (K.S.); 2120161037@bit.edu.cn (P.S.)

[2] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academic of Science, Beijing 100190, China; Yisheng.lv@ia.ac.cn

[3] School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260, USA

[*] Correspondence: zniu@bit.edu.cn; Tel.: +86-10-6891-5209

[†] These authors contributed equally to this work.

check for updates

Abstract: Traffic situation awareness and alerting assisted by adverse weather conditions contributes to improve traffic safety, disaster coping mechanisms, and route planning for government agencies, business sectors, and individual travelers. However, at the city level, the physical sensor-generated data are partly held by different transportation and meteorological departments, which causes problems of "isolated information" for data fusion. Furthermore, it makes traffic situation awareness and estimation challenging and ineffective. In this paper, we leverage the power of crowdsourcing knowledge in social media and propose a novel way to forecast and generate alerts for city-level traffic incidents based on a social approach rather than traditional physical approaches. Specifically, we first collect adverse weather topics and reports of traffic incidents from social media. Then, we extract temporal, spatial, and meteorological features as well as labeled traffic reaction values corresponding to the social media "heat" for each city. Afterwards, the regression and alerting model is proposed to estimate the city-level traffic situation and give the suggestion of warning levels. The experiments show that the proposed model equipped with gcForest achieves the best root mean square error (RMSE) and mean absolute percentage error (MAPE) score on the social traffic incidents test dataset. Moreover, we consider the news report as an objective measurement to flexibly validate the feasibility of proposed model from social cyberspace to physical space. Finally, a prototype system was deployed and applied to government agencies to provide an intuitive visualization solution as well as decision support assistance.

Keywords: city-level traffic alerting; adverse weather; social transportation; crowdsourcing knowledge; intelligent transportation system

1. Introduction

From the view of decision-making and disaster prevention, adverse weather (or inclement weather and extreme weather) is a significant factor affecting traffic [1]. For instance, an experiment showed that from 2005 to 2014, the number of traffic accidents caused by wind, snow, fog, dust, and hail equaled 287,783 in mainland China and resulted in 82,064 deaths [2]. Many studies have indicated that adverse weather negatively affects drivers' decision-making ability [3], the accident rate in traffic systems [4], and traffic flow [5]. Therefore, weather effects on transport systems are common research concerns among meteorological departments, transport agencies, and government administrations [6]. This work demands multichannel data collection, data extraction, knowledge fusion, environmental feature

introduction, and machine-leaning models to establish an intelligent transport system (ITS) for traffic situation estimation and alerting [7]. However, the physical sensor-based ITSs are time-consuming and expensive in deployment and maintenance, requiring large amounts of human power and resources [8]. The burden of data acquisition is also heavy for many government departments, especially in developing countries [9]. Furthermore, for historical and procedural reasons, transportation and meteorological departments in various regions and at various levels use different channels and systems to manage their business data, which results in significant obstacles to data aggregation and mining.

Today, the social sensor-based ITSs, or so-called social transportation, provide a significant opportunity for improving ITSs and thus have received increasing attention [10,11]. Social transportation collects, retrieves, and mines data from social media, GPS, and mobile phones by taking advantage of crowdsourcing, easy acquisition, and real-time data from the virtual world of the Internet and mobile communication [12]. The deployment of social transportation systems has demonstrated great power in traffic incident detection, traffic situation awareness, traffic flow forecasting, traffic opinion mining, routing plan, etc. [13]. However, how to build a social transportation system assisted by meteorological features is still uncertain. More specifically, although previous studies found that people's travel behavior can be affected by reading meteorologically related tweets [10], the relationship between meteorologically related opinions and the corresponding traffic reaction in social media still needs to be explored, which leaves an essential question unanswered: Can we build a city-level traffic situation awareness and alerting model assisted by adverse weather data that benefits from social media?

To answer this question, in this paper, by collecting temporal, spatial, traffic, and meteorological data, we mined the correlation between adverse weather topic heat and traffic incidents in social media, and further propose a traffic situation awareness and alerting model assisted by adverse weather data to provide information on city-level traffic situations reflected in social media. Finally, we verify these warning results from cyberspace with real-world urban traffic situations. To the best of our knowledge, this is the first time that traffic situations have been estimated and forecasted using social media assisted by adverse weather data. The road map of our research idea and approach is shown in Figure 1. In particular, we leverage the social path described in the figure below to achieve city-level traffic situation awareness and alerting. Moreover, by combining these social media data with physical sensed data, the ITSs can be shifted from cyber-physical systems (CPS) to cyber-physical-social systems (CPSS), which give more comprehensive and robust results for decision-making [11,14].
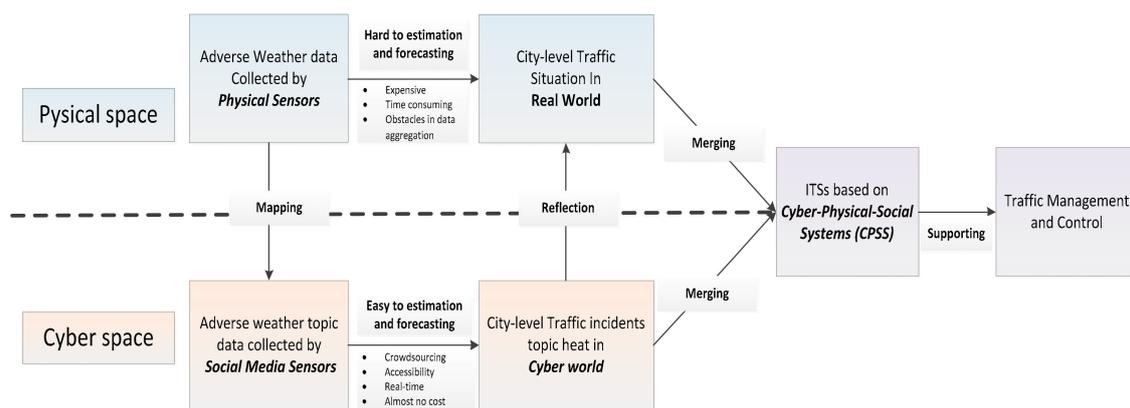

Figure 1. The road map of our research approach.

The rest of this paper is organized as follows. In Section 2, we give a review of related work from three aspects: the traffic situation awareness model in ITSs, social transportation in ITSs and meteorological research through social media. In Section 3, we present a city-level traffic situation awareness and alerting method assisted by adverse weather in social media, also giving the implementation details of the collection, extraction, regression, and alerting models. In Section 4,

we compare the results of regression models for traffic forecasting and flexibly validate the alerting results according to the reported traffic incident news. In Section 5, we demonstrate the prototype system of our proposed method, which has been applied by the China Meteorological Administration. In addition, other potential applications of our method are outlined. Finally, Section 6 concludes the paper and lists planned future work.

2. Related Work

2.1. Traffic Situation Awareness Models in ITSs

Previous research has presented several models of traffic situation awareness and alerting by constructing various kinds of ITSs. Many state-of-the-art models, including, but not limited to, deep learning [15] and wavelet-supported vector machines [16], have been utilized to estimate and report car crashes or accidents [17], roadside sulfur dioxide and nitrogen oxide data projections [18], identification of suspicious unlicensed taxis [19], the choice of transport mode [20], etc.

Significant work on ITSs combined with weather-related features has been performed. Dey et al. [21] reviewed the influence of adverse weather on ITSs before 2015, proving that the research in this area had great potential for investigation. Park et al. [22] summarized the features of accidents including adverse weather on the expressway to Busan, Korea, in addition to building a detection model based on videos and other sensors deployed on the expressway. Lee et al. [23] made full use of the weather data along the freeway from Seoul to Gyeongpodae to assess levels of traffic congestion using multiple linear regressions. Tomas et al. [24] explored a scheme and system architecture to deploy forecasting models assisted by adverse weather in a real road network and discussed an efficient way to provide early warnings. Stamos et al. [25] explored the impact of adverse weather on traffic from a data-driven perspective. Yu et al. [26] linked the reason for freeway accidents to weather from a micro perspective. As to machine learning methods, Tsirigotis et al. [27] selected the vector and Bayesian methods to estimate changes in traffic flow over a short period of time caused by adverse weather. With the development of deep learning methods, Koesdwiady et al. [28] chose a DBN (deep belief network) model with the input parameter of weather data and made the model reduce forecasting error.

In terms of climate, Keay et al. [5] provided an analysis of traffic safety issues in terms of climate change in Canada. Koetse et al. [29] investigated similar issues of the impacts on transport from the perspective of climate change rather than meteorological analysis. All the ITS models above provide a solid basis to determine the method and application for weather-affected traffic safety alerting, traffic flow control, and travel choice optimization. However, there are no works that map the relationships between weather and traffic situations in social cyberspace.

2.2. Social Transportation in ITSs

Social media, in the form of crowdsourcing, offers advantages in terms of real-time feedback and multiple information sources, which play an essential role in the integration and dissemination of information. The social transportation has been becoming a new research trend in ITSs for traffic management and control [10,12,13,30]. Problems that are rarely solved in the physical world can be addressed in cyberspace with the social transportation approach [31].

For example, we can use Global Positioning System (GPS) data from taxis to measure traffic congestion based on speed [32]. Maghrebi et al. [33] extracted tweets through Twitter Stream to analyze how people travel in their daily lives and found that walking and driving were the most frequent travel modes. Anantharam et al. [34] implemented the conditional random field (CRF) method to annotate entities of traffic event from Twitter. In another application, Ni et al. [35] forecasted metro passenger flow through aperiodic social media data combined with periodic flow data in the Mets–Willets Point metro station, New York. A similar forecasting was also proposed from a more precise geometric granularity [36].

Opinion mining technologies also have been used in social transport systems. Zeng et al. [37] investigated the characteristics of traffic congestion in social media considering the significant traffic in China and evaluated the opinions of popular topics among Chinese "netizens" (Internet users). Cao et al. [38] proposed the traffic sentiment analysis (TSA) tool, which takes advantage of rule- and learning-based approaches to process traffic web data. The method and tool conducted opinion analysis on the "yellow light rule" and "fuel price" traffic incidents in China. As noted, to the best of our knowledge, little research has combined meteorological and traffic social data to estimate the reflection of the real-world traffic situation.

2.3. Meteorological Research through Social Media

Although research on the relationship between meteorological topic and transportation situation alerting in social media has not been performed, the social-media-based behavior analysis under different weather conditions has attracted more and more attention [39–41].

Kirilenko and Stepchenkova [42] collected and visualized tweets about global climate change in 2014 in four languages. They pointed out that, although the number of messages related to climate change is significant, the flow of information is highly inactive. They also pointed out that very few media outlets, celebrities, and highlighted bloggers are leading the debate. In the early warning field, Grasso et al. [43] used Twitter hashtags as a classification label to apply meteorological social opinion analysis in weather forecast system. They regard every hashtag as a weather event and create a statistical experiment using these social-based weather events.

Other studies tend to start from the psychology or cognitive science point of view. Park et al. [44] compared the data of temperature, humidity, and atmospheric pressure with tweets and found that temperature and pressure were correlated positively to the number of positive emotional texts, whereas humidity was correlated negatively to the number of negative emotional texts. Li et al. [45] conducted a more systematic and in-depth investigation, observing that mood varies with weather conditions like temperature and pressure. An et al. [46] investigated the emotional issues related to climate change in social media, and Cody et al. [47] conducted a more in-depth investigation based on [46], indicating that the climate bill and oil drilling content would reduce happiness, whereas climate rallies, book distribution, and green ideological competitions could increase happiness. These works show that social media is a valuable resource for reflecting weather and climate effects on people's emotions and behaviors.

3. Methods and Modeling

At the road level, substantial empirical research has shown that physical data are related to highway congestion and flight delays [31]. However, at the city level, the data-driven model is ineffective because of the problems of "isolated information" and the difficulty of fusing multi-source data in a timely manner. Faced with city-level traffic awareness and alerting problems, we propose a model to extract adverse weather- and traffic-related tweets from Sina Weibo, a microblogging service for Chinese-speaking people, to investigate the social reaction on adverse weather topics and traffic incidents. The urban traffic situation is perceived and estimated by weather topic heat, adverse weather types, temporal and spatial features, etc., and then the alerting model is built by measuring the properties and dimensions of traffic incidents.

The architecture of the system we propose is shown in Figure 2. We filtered tweets related to topics from Weibo and stored them in our database to make sure that each tweet contains either an adverse weather topic or a traffic topic. Then, the preprocess submodule conducted word segmentation and other processing work, because Chinese words are separated based on semantic implementation rather thanspaces like English. After tweets (or records) are preprocessed, we extracted properties like location, adverse weather keywords, traffic keywords, published time, etc., and aggregated all tweets into adverse-weather-influenced traffic incidents in social media. Afterwards, we calculated the heat of each traffic incident and the adverse-weather-related topic, then built the dataset for traffic

awareness and alerting. The results of this analysis could also be used to enrich services in many applications. Finally, domain experts review the system results for further evaluation, investigation, and decision-making.
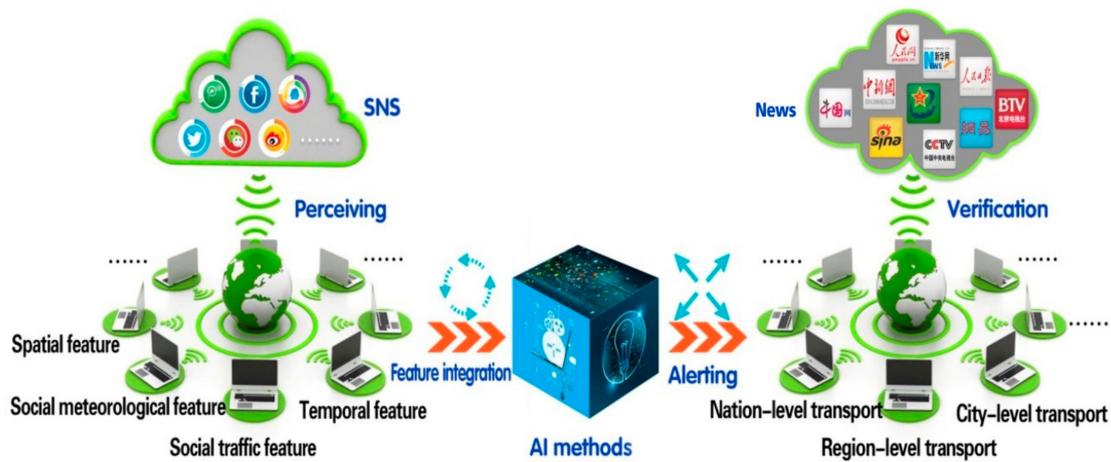

Figure 2. Architecture of city-level traffic situation awareness and alerting model.

3.1. Data Collection with Word2vec-Based Social Sensors

In this paper, we built a micro-blogging-based traffic incident dataset that contains the influence factor of adverse weather from a particular location within a certain period of time. We use the principle of the "3W" dimensions (i.e., "When, Where and What" attributes). Specifically, we regarded time, location, and incidents as the three elements from social media that could also be inferred from a physical event. To begin collecting data from Weibo, we needed a keyword list for the crawler to detect related tweets.

In terms of keywords in the traffic domain, to the best of our knowledge, few systemic or widely accepted keyword lists describe traffic incidents (or "event categories"). Therefore, we constructed a keyword list in the traffic domain by using a statistical method.

Word2vec model (https://code.google.com/archive/p/word2vec/) creates a vector space (word embedding) by using artificial neutral networks to retain the words' original linguistic meaning. The word embedding models make it possible to measure the semantic similarity between words, and predict which words with similar semantic meaning will be close to each other in the vector space [48,49].

Assuming X and Y are two vectors with n dimensions and we let $x_i \in X$ and $y_i \in Y$, the cosine similarity can be calculated as follows:

$$Sim(X,Y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}} \tag{1}$$

Unlike English sentences, Chinese sentences are separated semantically rather than by spaces, so all Chinese text needs to be segmented before using Word2vec. In this paper, we conducted word segmentation using Python packages from Jieba (https://github.com/fxsjy/jieba), and then trained a Word2vec model with 200 dimensions using 1,049,823 traffic-related tweets and 324,130 traffic-related news records from June 2016 to June 2017.

By combining news and tweets data into a training corpus, the wording styles of casual expression in tweets and formal expression in news can both be learned by word2vec, which provides more precise word embedding results for message filtering and keyword extension. An example of word embedding for traffic is given in Table 1, which shows the closest words for the traffic keywords "traffic congestion" in our trained Word2vec model.

Here, a threshold of 0.5 is set in the word2vec model to find synonyms (extended keywords), and all the synonyms were manually double checked by the authors and domain experts. Finally, we built the traffic keywords list with traffic seed keywords and extended keywords from traffic word embedding. The complete keyword list is given in Table 2 in Chinese; note that we translate all keywords into English for better comprehension.

In terms of keywords for adverse weather, we applied Chinese National Standard GB/T 27962–2011 into our categories for 14 types of adverse weather. Based on these 14 kinds of adverse weather, we first used word2vec to find similar words, as done for the traffic keyword list, and then had domain experts double check them to select key words. Table 3 gives a brief introduction as well as the final weather-related keyword list we used in our experiment; note that some different Chinese words correspond to the same English translations.

Table 1. Most similar words for "traffic congestion".

Words	Similarity
堵塞 (jam)	0.952356529236
堵情 (congestion situation)	0.898310112953
交通阻塞 (traffic jam)	0.893572020531
堵车 (heavy traffic)	0.80860207081
路阻 (road block)	0.740446996689
空气污染 (air contamination)	0.680577373505
停车难 (difficulty of parking)	0.678475296497
内涝 (waterlogging)	0.655691361427
供需矛盾 (contradiction between supply and demand)	0.6066395998
停运 (stoppage in transit)	0.598671365489

Table 2. The final traffic keyword list.

Type	Keywords
congestion related	交通拥堵 (traffic congestion), 交通堵塞 (traffic jam), 交通阻塞 (traffic block), 交通停运 (traffic halt), 交通受阻 (disrupted transportation), 交通中断 (interruption of transport communication), 交通管制 (traffic control), 交通限流 (traffic flow limiting), 交通瘫痪 (traffic paralysis), 堵车 (traffic jam), 高速关闭 (highway closing), 高速封闭 (closed highway), 高速中断 (highway interruption), 高速堵塞 (highway congestion), 高速阻塞 (highway block), 高速拥堵 (highway congestion), 高速管制 (highway control), 高速限行 (highway flow control), 高速停运 (highway halt), 早高峰拥堵 (morning rush hour), 晚高峰拥堵 (evening rush hour), 道路封闭 (road closure), 路段封闭 (road section closed)
accident related	交通事故 (traffic accident), 高速事故 (highway accident), 车祸 (car crash), 撞车追尾 (shunt), 车辆剐蹭 (vehicle rub)
long-distance transport related	航班延误 (flight delay), 航班取消 (flight cancellation), 航班停运 (flight outage), 列车晚点 (train delay), 列车停运 (train halt), 停航 (suspension of ships)

Table 3. Fourteen types of inclement weather and the final weather keyword list.

Category	Warning Levels	Example words	
台风 (Typhoon)	I, II, III and IV	台风 (typhoon), 热带风暴 (tropical storm), 热带气旋 (tropical cyclone), 飓风 (hurricane)	
暴雨 (Rain Storm)	I, II, III and IV	暴雨 (rain storm), 暴风雨 (tempest), 强降雨 (heavy rainfall), 强降水 (severe precipitation)	

Table 3. *Cont.*

Category	Warning Levels	Example words	
暴雪 (Snow Storm)	I, II, III and IV	暴雪 (snow storm), 暴风雪 (blizzard), 雪暴 (buran)	
寒潮 (Cold Wave)	I, II, III and IV	寒潮 (cold wave), 寒流 (cold surge)	
大风 (Gale)	I, II, III and IV	大风 (gale), 狂风 (fierce wind), 强风 (wild wind), 暴风 (fierce wind)	
沙尘暴 (Sand Storm)	I, II and III	沙尘暴 (sand storm), 沙暴 (desert storm), 黑尘暴 (dust storm)	
高温 (Heat Wave)	I, II and III	高温 (heat wave), 热浪 (high temperature), 桑拿天 (suana weather), 酷暑 (canicule), 三伏天 (dog days), 炎热 (blazing)	
干旱 (Drought)	I and II	干旱 (drought), 旱灾 (aridity)	
雷击 (Lightning)	I, II and III	雷击 (lighting), 雷暴 (thunder), 打雷 (thunder), 雷电 (thunder and lightning)	
冰雹 (Hail)	I and II	冰雹 (hail), 降雹 (hail fall), 风雹 (hail fall), 雹灾 (hailstorm)	
霜冻 (Frost)	II, III and IV	霜冻 (frost), 霜降 (frost descent)	
大雾 (Heavy Fog)	I, II and III	大雾 (heavy fog), 浓雾 (thick fog), 雾灾 (mist)	
霾 (Haze)	II and III	霾 (haze), 雾霾 (smog)	
道路结冰 (Road Icing)	I, II and III	道路结冰 (road icing), 路面结冰 (icy road), 公路结冰 (road icing)	

3.2. Data Filtering and Feature Extraction

On the basis of collected data and obtained keyword lists (including synonyms), the rule-based filtering strategy were utilized. Specifically, all the messages that either contain the words in the traffic keyword list or the words in the adverse weather keyword list were regarded as the related messages. Moreover, all tweet authors with fewer than 10 followers were regarded as spammers, and their corresponding messages were also filtered out as spam (false data).

The data relevance has been ensured through the data collection and filtering steps. Then, the temporal features (date, season, holiday, weekday, etc.), location features (longitude, latitude, urban, countryside, etc.) and the meteorological features (including social features, such as different focus rate of each type of adverse weather, and also the real meteorological data like temperature, humidity

and sea level pressure) that may potentially influence the traffic situation in both the real world and cyberspace were extracted.

First, the temporal features were directly obtained from the tweets and converted to a timestamp format, and then mapped to other temporal features, such as season, holiday, and weekday. Second, we analyzed the contents of the tweets to resolve location features [34]. The locations were extracted from one tweet according to word frequency and looked up in the place name database to transfer the information to the prefecture-level city to which it belonged [50]. Third, the adverse weather types and traffic incidents were extracted based on the abovementioned word list. Finally, based on traffic word embedding, we aggregated all the tweets into city-level traffic incidents with retweet number as their heat or focus rate [51].

After data collection and preprocessing, the features and labels are obtained, and thus we can estimate the traffic situation for cities and countries through the process of "adverse weather tweets → temporal, spatial, meteorological features → traffic regression models → traffic incidents heat → alerting models → traffic warning level." The regression model and alerting model will be discussed in the next section.

3.3. Traffic Regression Models

Next, the Weibo dataset that contains traffic incidents or adverse weather messages, as well as the three typical feature sets of time, location, and adverse weather were built. Specifically, we regard the heat value of each traffic incident as the forecasting label. To find the relevance between adverse weather and traffic incidents in social media, and make an estimation of city-level traffic incident heat for the same period in the same city, we vectored all the features and used a regression model to achieve that goal.

The regression model determines a function and corresponding parameters that can accurately forecast future city-level traffic incident heat. To ensure ideal regression performance, we selected an appropriate model by comparing and evaluating the models of a Gradient-Boosting Regression Tree (GBRT) [52], Random Forest Regression (RFR) [53], Support Vector Regression (SVR) [54] and Linear Regression (LR) [55]. We implemented the above regression work by calling APIs from packages in Scikit-Learn [56,57]. To benefit from the strength of deep learning technologies, we also selected Stacked Auto-Encoder (SAE) [14] and gcForest [58] as deep neural network and deep random forest to improve the precision of regression.

Note that, facing the real-time city-level traffic incidents forecasting problem in this paper, we only employed social media data for traffic regression because the messages from social media are more timely than the messages from news media. Although the messages from both news media and social media can be involved in the regression model for better accuracy, the timeliness of forecasting model could be reduced. Moreover, additional research about the combination of messages from news media and social media still needs to be performed, such as processing with different text length, calculation of cross-platform incidents' heat, weight setting of multi-source messages, etc.

3.4. Traffic Alerting Model

The traffic alerting model mapped a regression result to a certain warning level. We defined H_{Min} and H_{Max} equal to the maximum and minimum values of the forecasting traffic incident heat in a normative data sample, respectively. All other values were mapped within the range of 0–1 based on this maximum and minimum value as follows:

$$H_i' = \frac{H_i - H_{\text{Min}}}{H_{\text{Max}} - H_{\text{Min}}} \tag{2}$$

Note that H_i represents each traffic incident heat and H_i' represents the normalized traffic incident heat.

The traffic incidents at the highest level of early warning rarely occur. If we divided the level equally, the highest warning level will be easily triggered. To reflect differences among these warning levels, we were inspired by the JCR Journal Partition designed by the National Science Library, Chinese Academy of Sciences (http://www.fenqubiao.com) for which the top academic journals in the first level only account for 5%. The warning levels are defined as follows.

First, the total number of forecasted city-level traffic incidents was defined as n, and all their forecasting heat values were ranked in reverse order. The traffic incidents with top 5% forecasting heat values were regarded as level 1 warning incidents. The heat value threshold of level 1 warning incidents was defined as H_{t_1} where $t_1 = [n \times 0.05]$. Then let S equal the total number of heat values for the remaining 95% of traffic incidents, which is represented as the following formula:

$$S = \sum_{i=t_1+1}^{n} H_i' \tag{3}$$

We then checked the remaining 95% traffic incidents' heat values, denoting the heat values interval of level 2 warning incidents as $[H_{t_2}, H_{t_1})$. Starting from an index of $t_1 + 1$, the reverse-ranked traffic heat values were added one by one until the cumulative value was greater than or equal to S/3. The low bonder index of level 2 warning interval was defined as t_2, which can be derived as follows:

$$t_2 = \underset{m}{\arg\min} \left(\sum_{j=t_1+1}^{m} H_j' \geq S/3 \right) \tag{4}$$

The low bonder index of level 3 warning interval was denoted as t_3, which can be obtained in the same way of deriving t_2. Thus, when a new incident heat value h arrives, we derive the warning level from the following equation:

$$Level\ (h)\ =\ \begin{cases} 1, & h \geq H_{t_1} \\ 2, & H_{t_1} > h \geq H_{t_2} \\ 3, & H_{t_2} > h \geq H_{t_a} \\ 4, & \text{otherwise} \end{cases} \tag{5}$$

The application of this warning level partition could make different warning level variations within incidents. For better comprehension, we randomly sampled 3200 incidents from the total number of incidents and their warning level distribution is shown in Figure 3.
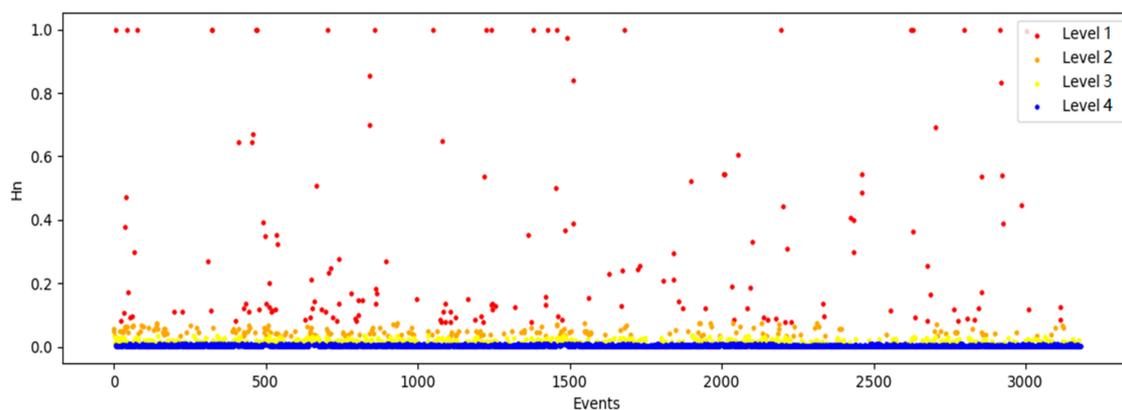

Figure 3. Warning level distribution of 3200 samples.

4. Experiments

We collected adverse-weather-affected traffic data from Weibo, preprocessed it as discussed in Section 3, and finally obtained a total of 128,815 tweets from 1 January to 1 August 2017 as the dataset.

The dataset was separated into training data, validation data, and test data. The details are shown in Table 4. For the city selection, we selected the 31 capitals of each province, autonomous region, and municipality in mainland China as monitoring cities. Furthermore, to verify whether the traffic incidents in social media reflected the real-world traffic situation, we also collected 15,130 news items from 1 January to 1 August 2017, which is discussed below.

Table 4. Dataset description.

Training Dataset		Validation Dataset		Test Dataset		Physical Space: Verification Dataset
Weibo Data	Traffic Incidents	Weibo Data	Traffic Incidents	Weibo Data	Traffic Incidents	News Data
128,815	28,024	10,993	1830	1213	200	15,130

In the following experiments, we first tried six types of regression models to verify whether there are relationships between adverse weather and traffic incidents in social media; we also selected a model for city-level warning according to the model evaluation results. Second, we mapped and visualized the warning levels of 200 traffic incidents. Finally, we verified the relationship of traffic incidents in social media to real-world traffic incidents to further prove that our proposed city-level traffic situation awareness and alerting model can be deployed in both cyberspace and physical space.

4.1. Experiments on Traffic Regression and Alerting

According to the dataset mentioned above, we measured the performance of the models by using two indicators: mean absolute percentage error (MAPE) and root mean square error (RMSE). Assuming there are n forecasting values and labels, the MAPE can be formulated as follows:

$$E_{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y'_i - y_i}{y_i} \right| \tag{6}$$

where y'_i is a forecasting value and y is the true label value. The RMSE can be formulated as:

$$E_{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y'_i - y_i)^2}{n}} \tag{7}$$

Moreover, to test the performance of the regression model, we selected 200 traffic incidents from 0:00 21 July to 24:00 31 July manually, and compared the top-10 forecasted traffic incidents, ordered by their heat. The forecasting results include location, time (with period), and traffic incident heat. The real traffic situations were manually marked by searching for news items in our database. In the experiment, the structure of the stacked auto-encoder is the same as in [58] (except for the input format). In addition, the structure of gcForest is a four- to 20-layer model, with each layer consisting of two random forests as well as two gradient tree boosting trees in each layer. As shown in Table 5, the deep models outperform other machine learning algorithms and the state-of-the-art deep random forest has the best performance.

Table 5. Comparison of regression models' performance on the test set.

Model	MAPE	RMSE
Linear Regression (LR)	496.89	50.83
Support Vector Regression (SVR)	355.67	22.25
Random Forest Regression (RFR)	263.10	21.00
Gradient Boosting Regression Tree (GBRT)	252.09	20.41
Stacked auto encoder (SAE)	239.18	20.05
gcForest (Deep Forest)	168.08	18.97

In addition, we successfully forecasted nine out of 10 traffic incidents, as shown in Table 6, proving that our social media approach is feasible.

Based on the 200 traffic incident forecasting values, we mapped the values into warning levels with the proposed model. We also compared the distribution with the real warning level, as shown in Figure 4. The similar trends demonstrate that our approach achieves the expected effect.
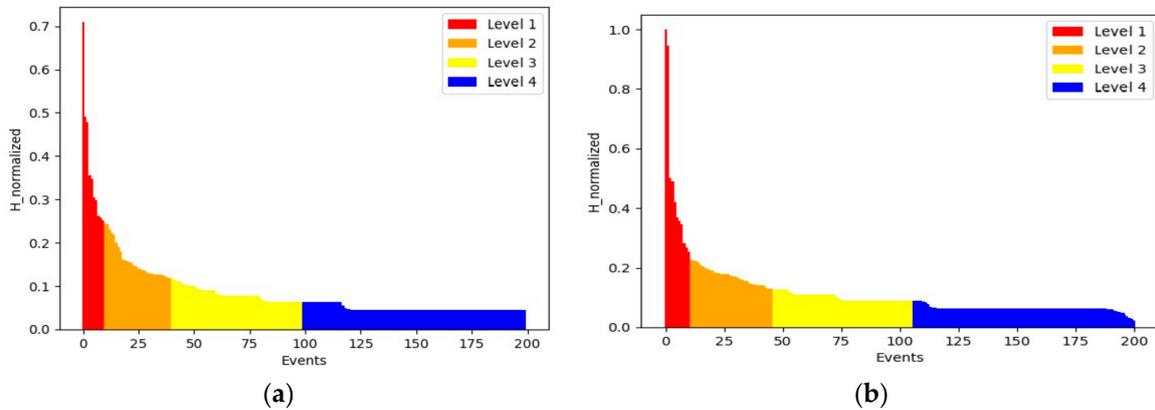
Figure 4. Comparison of (**a**) alerting warning level distribution and (**b**) labeled warning level distribution.

Table 6. Comparison of top 10 social traffic events in forecasting and labeled traffic situation.

Rank	Forecasting Event	Forecasting Heat Value	Rank	Labeled Event	Labeled Heat Value	Labeled Traffic Situation
1	Shijiazhuang 21 July 12: 00–18: 00	120.13	1	Shijiazhuang 21 July 12: 00–18: 00	497.0	traffic interruption, flight delay
2	Kunming 20 July 6: 00–12: 00	113.61	2	Kunming 20 July 6: 00–12: 00	224.0	traffic interruption, highway interruption
3	Beijing 28 July 6: 00–12: 00	45.98	3	Beijing 28 July 6: 00–12: 00	63.0	traffic congestion
4	Chengdu 28 July 6: 00–12: 00	32.77	4	Guangzhou 21 July 18: 00–24: 00	60.0	road closure
5	Shanghai 27 July 6: 00–12: 00	29.24	5	Changsha 25 July 6: 00–12: 00	44.0	road closure
6	Taiyuan 24 July 6: 00–12: 00	26.37	6	Chengdu 28 July 6: 00–12: 00	34.0	flight delay
7	Beijing 28 July 12: 00–18: 00	24.62	7	Shanghai 27 July 6: 00–12: 00	30.0	traffic congestion
8	Changsha 25 July 6: 00–12: 00	23.91	8	Beijing 28 July 12: 00–18: 00	30.0	traffic congestion
9	Shanghai 24 July 6: 00–12: 00	11.80	9	Taiyuan 24 July 6: 00–12: 00	20.0	traffic interruption
10	Haikou 24 July 12: 00–18: 00	7.80	10	Haikou 24 July 12: 00–18: 00	18.0	traffic control

4.2. From Cyberspace to Physical Space: Flexible Verification

Although we used a machine-learning method to learn and simulate the relationship between adverse weather topic heat and traffic incidents in social media, we need to answer an additional question to demonstrate the validity of the proposed city-level traffic awareness and alerting model: Does this model reflect real-world situations? In other words, if we detect and track one "hot traffic incident" in social media, does this event exist in real life?

Owing to the objectivity and authenticity of news reports, we regarded traffic news reports as objective measurements that indicate real-world hot traffic incidents. Specifically, we assume that all traffic hot incidents are reported in the news. Therefore, we deployed a web crawler to collect the news report data about traffic incidents that took place in Beijing, China between 1 January and 31 July 2017 as a verification dataset.

Next, we use the regression models to forecast the heat value of traffic incidents in Beijing. For better visualization, we divided the whole validation time period into time blocks, with each time block containing six hours. We draw all time blocks whose forecasted heat value is larger than zero (i.e., traffic incidents) as in Figure 5. Note that H represents the forecasted heat of each time block. Then, we applied our alerting model to map the heat value into traffic warning levels; seven level 1 warning incidents were marked with Arabic numbers in Figure 5. It is worth mentioning that the heat values of the seven alerted traffic incidents are all over 500.

Moreover, to gain a more intuitive understanding, we tracked seven level 1 warning traffic incidents with a heat value above 500 and compared them with the traffic incidents reported in the news in the verification dataset in Table 7. Here, we define all traffic incidents with a warning level of 1 as forecasted "hot traffic incidents."
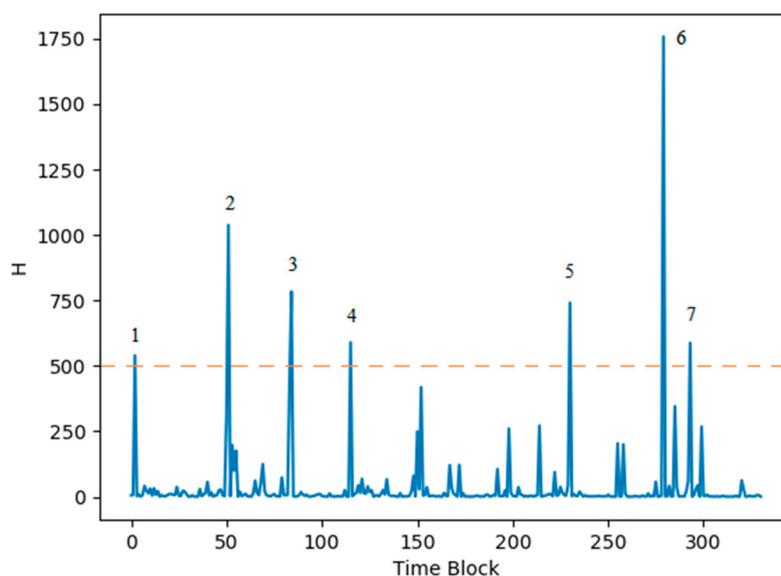

Figure 5. Variation of social traffic heat in Beijing from 1 January to 31 July 2017.

Table 7. Social traffic incidents with heat value above 500 in Beijing from 1 January to 31 July 2017 and corresponding news reports.

No.	Date	Time Period	Weekday	Heat Value	Corresponding Heat of Weather (Top 3)	Warning Level	Corresponding News (Facts)
1	2 January	6:00–12:00	Monday	540.0	Heavy fog: 52 Frost: 2 Haze: 1	1	新华网：北京遇大雾红色预警多条高速采取封闭措施 (Xinhua Net: Beijing met red alert of heavy fog and a number of high-speed closed)
2	16 January	18:00–24:00	Monday	1038.0	Haze: 48 Snow storm: 3 Gale: 2	1	千龙网：北京:雾霾袭扰多条高速路封闭 (Qianlong Net: Beijing: haze attack numbers of high-speed road which result in closing)

Table 7. *Cont.*

No.	Date	Time Period	Weekday	Heat Value	Corresponding Heat of Weather (Top 3)	Warning Level	Corresponding News (Facts)
3	28 January	12:00–18:00	Saturday	784.0	Haze: 40 Gale: 26 Heavy fog: 2	1	北京晚报：大风救驾大年初二:北京大风蓝色预警雾霾再见 (Beijing Evening News: the wind saves the Lunar New Year's Day: Beijing says goodbye to blue warning of haze) 凤凰网：北京启动"空城"模式人都去哪了? (Phoenix Net: Beijing start "empty city" model:where are the people?)
4	11 February	6:00–12:00	Saturday	590.0	Snow storm: 101 Haze: 95 Cold wave: 6	1	网易新闻：北京喜迎首场春雪:部分地区雪量较大，影响交通出行 (Netease News: Beijing in the first spring snow: huge snow volume in some area affected traffic and travel)
5	21 March	12:00–18:00	Tuesday	741.0	Heavy fog: 182 Haze: 165 Sand Storm: 9	1	中国青年网：北京遭遇雾霾天，使京藏高速进京方向发生车祸造成大面积拥堵 (China Youth Network: Beijing encountered haze days, Beijing–Tibet high-speed has a large area of traffic congestion caused by car accidents)
6	7 June	6:00–12:00	Wednesday	1758.0	Rain storm: 247 Gale: 86 Heavy fog: 34	1	网易新闻：2017高考未破"下雨魔咒"高峰路况拥堵 (Netease News: 2017 college entrance examination obey "the rain curse", congestion worsen by the weather during the morning peak)
7	4 July	6:00–12:00	Tuesday	588.0	Rain storm:461 Heat wave: 40 Haze:15	1	网易新闻：北京首都机场受雷雨天气影响已取消航班113架次 (Netease News: Beijing Capital Airport has canceled 113 flights due to a thunderstorm)

The results show that all the hot traffic incidents flagged by our social media approach were also reported by the news media, which means the traffic incidents noted in social media really happened in the real world. Furthermore, we verified the model-alerted traffic incidents with lower warning levels, and rechecked the news-reported hot traffic incidents in the verification dataset. Our proposed model reported seven traffic incidents and missed four. For the convenience of understanding, the confusion matrix is shown in Table 8. The precision, recall, and F1-score are 1.0, 0.636, and 0.777, respectively. The preliminary verification proved that there is a correlation between social cyberspace and physical space.

Table 8. Confusion matrix of proposed alerting model in Beijing.

		Actual Incidents	
		Hot Traffic Incident (Warning Level 1)	Non-Hot Traffic Incident
Forecasting incidents	Hot traffic incident (warning level 1)	7	0
	Non-hot traffic incident	4	335

5. Prototype and Potential Applications

Based on the previous method, we developed a prototype system called the adverse-weather-affected traffic incidents perception and warning system to visualize the collected data, extract features, statistics results, and alerting incidents, as shown in Figure 6.

Figure 6. Weather-related traffic incidents perception and warning system.

The prototype system has been initially deployed in CMA. The proposed method demonstrated advantages of feasibility, timeliness, and high accuracy compared to the traditional physical methods. Both the model and the system are still iteratively improving and upgrading according to CMA's management and decision requirements. In addition, the prototype system provides a consistent service for CMA, including adverse-weather-affected city-level traffic situation alerting and adverse weather impacts on traffic situations. In particular, the prototype system successfully forecasted the serious traffic situation caused by typhoons Hato and Pakhar in the south of China in mid-July 2017. Our proposed social approach will also be replicated to support the long-term plan, is the so-called "Meteorology Plus Plan" of the CMA. The plan includes "Meteorology + Agriculture," "Meteorology + Emergency Management," "Meteorology + Tourism," "Meteorology + Disease Prevention," "Meteorology + urban planning," etc.

In addition to the meteorological administration departments, the potential applications of our proposed method have also attracted attention from transportation administration deportments like Qingdao municipal commission of transport, and the preliminary intention of cooperation has been reached. Weather is closely related to people's productiveness and activities. The potential value of our proposed method still needs to be explored. We believe that, with further development and improvement, the prototype could be widely applied by governments, businesses, and individuals in fields that are highly correlated with meteorology like agriculture, tourism, and health care.

6. Discussion and Conclusions

Owing to the time-consuming deployment, high-cost maintenance, and isolated-information problem in physical-sensor-based ITSs, a social-sensor-based ITS for adverse-weather-affected city traffic awareness and warning was proposed in this paper. The proposed ITS leverages the advantages of social media, such as crowdsourcing, real time, multiple sources, and easy acquisition. The social approach has been validated experimentally in cyberspace and physical space.

Specifically, for data collection, the word2vec for transportation was trained to construct the traffic search keyword list. In addition, China National Standard GB/T 27962-2011, which contains 14 types of adverse weather conditions, was used to construct the meteorology search keyword list. Combining these two word lists, we collected 128,815 tweets from 1 January to 1 August 2017. For data preprocessing, three types of features that influence the traffic incidents—temporal, spatial, and meteorological features—were extracted. Meanwhile, the traffic incidents' heat values were also calculated as the forecasting target label. For forecasting and alerting, we compared the effectiveness of six typical machine learning models of traffic incident forecasting. For verification, we considered news media as an objective measurement and took Beijing as a case study to prove the generality of our proposed social approach in the real world.

Although our proposed model is a preliminary attempt at city-level traffic situation awareness and alerting assisted by adverse weather in a social approach, additional valuable work can be conducted in the future. First, the authority of each related tweet should be considered. The authors may be categorized as governments, news agencies, companies, spammers, online sellers, Internet water army and Internet celebrities, etc. and the posted tweets may be categorized as news abstracts, announcements, records, review, spams, advertisements, fake news, etc. Different types of authors and tweets should be given unequal weight in the forecasting model to achieve higher accuracy [59]. Second, the model needs to consider more impacts that may be related to traffic incidents, e.g., geological disasters [60,61], disease disasters, social events, etc. Integration of different types of impacts into a unified model will provide much better support for decision-making. Third, instead of machine learning-based approaches, this problem could also benefit from knowledge-based decision support methods. The knowledge services of transport and meteorology are both important to relieve the pinch points of social management. Finally, when improving and validating the proposed methods in various domains related to meteorology, more potential values need to be explored by the government and the public in the near future.

Author Contributions: Conceptualization, H.L.; Data analysis, Y.Z. and K.S.; Methodology, H.L., Y.Z. and K.S.; Project administration, Z.N.; prototype design, all authors; Supervision and suggestion, Y.L. and Z.N.; Visualization, P.S.; Writing, H.L. and Y.Z.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mahmassani, H.S.; Dong, J.; Kim, J.; Chen, R.B.; Park, B. *Incorporating Weather Impacts in Traffic Estimation and Prediction Systems*; US Department of Transport: Washington, DC, USA, 2009; Volume 108.

2. Ning, G.; Kang, C.; Chen, D.; Sun, G.; Liu, J.; Wang, S.; Shang, K.; Ma, M. Analysis of characteristics of traffic accidents under adverse weather conditions in china during 2005–2014. *J. Arid Meteorol.* **2016**, *34*, 753–762.

3. Zhang, L.; Colyar, J.; Pisano, P.; Holm, P. Identifying and assessing key weather-related parameters and their impact on traffic operations using simulation. In *Compendiums of 2002 Institute of Transportation Engineer Annual Meeting. CD-ROM. Institute of Transportation Engineers*; Citeseer: State College, PA, USA, 2003.

4. Brodsky, H.; Hakkert, A.S. Risk of a road accident in rainy weather. *Accid. Anal. Prev.* **1988**, *20*, 161–176. [CrossRef]

5. Keay, K.; Simmonds, I. The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accid. Anal. Prev.* **2005**, *37*, 109–124. [CrossRef] [PubMed]

6. Hranac, R.; Sterzin, E.; Krechmer, D.; Rakha, H.A.; Farzaneh, M.; Arafeh, M. *Empirical Studies on Traffic Flow in Inclement Weather*; Virginia Tech Transportation Institute: Blacksburg, VA, USA, 2006.

7. Amin, M.S.R.; Zareie, A.; Amador-Jimenez, L.E. Climate change modeling and the weather-related road accidents in Canada. *Transp. Res. Part D Transp. Environ.* **2014**, *32*, 171–183. [CrossRef]

8. Jing, Q.; Vasilakos, A.V.; Wan, J.; Lu, J.; Qiu, D. Security of the internet of things: Perspectives and challenges. *Wirel. Netw.* **2014**, *20*, 2481–2501. [CrossRef]

9. Zhai, Y.; Li, X. Advances in traffic meteorological service under the influence of disastrous weather. *J. Catastrophol.* **2015**, *30*, 144–147.

10. Zheng, X.; Chen, W.; Wang, P.; Shen, D.; Chen, S.; Wang, X.; Zhang, Q.; Yang, L. Big Data for Social Transportation. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 620–630. [CrossRef]

11. Wang, F. Scanning the Issue and Beyond: Real-Time Social Transportation with Online Social Signals. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 909–914. [CrossRef]

12. Wang, F.Y.; Zhang, J.J. Transportation 5.0 in CPSS: Towards ACP-based society-centered intelligent transportation. In Proceedings of the 2017 IEEE International Conference on Intelligent Transportation Systems, Yokohama, Japan, 16–19 October 2017; pp. 762–767.

13. Lv, Y.; Chen, Y.; Zhang, X.; Duan, Y.; Li, N. Social Media Based Transportation Research: The State of the Work and the Networking. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 19–26. [CrossRef]

14. Xiong, G.; Zhu, F.; Liu, X.; Dong, X.; Huang, W.; Chen, S.; Zhan, K. Cyber-physical-social System in Intelligent Transportation. *IEEE/CAA J. Autom. Sin.* **2015**, *2*, 320–333.

15. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 865–873. [CrossRef]

16. Sun, Y.; Leng, B.; Guan, W. A novel wavelet-svm short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [CrossRef]

17. Abdel-Aty, M.A.; Pemmanaboina, R. Calibrating a real-time traffic crash-prediction model using archived weather and its traffic data. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 167–174. [CrossRef]

18. Zito, P.; Chen, H.; Bell, M.C. Predicting real-time roadside CO and NO_2 concentrations using neural networks. *IEEE Trans. Intell. Transp. Syst.* **2008**, *9*, 514–522. [CrossRef]

19. Yuan, W.; Deng, P.; Taleb, T.; Wan, J.; Bi, C. An unlicensed taxi identification model based on big data analysis. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1703–1713. [CrossRef]

20. Anta, J.; Pérez-López, J.B.; Martínez-Pardo, A.; Novales, M.; Orro, A. Influence of the weather on mode choice in corridors with time-varying congestion: A mixed data study. *Transportation* **2016**, *43*, 337–355. [CrossRef]

21. Dey, K.C.; Mishra, A.; Chowdhury, M. Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1107–1119. [CrossRef]

22. Park, S.-H.; Kim, S.-M.; Ha, Y.-G. Highway traffic accident prediction using VDS big data analysis. *J. Supercomput.* **2016**, *72*, 2815–2831. [CrossRef]

23. Lee, J.; Hong, B.; Lee, K.; Jang, Y.-J. A prediction model of traffic congestion using weather data. In Proceedings of the 2015 IEEE International Conference on Data Science and Data Intensive Systems (DSDIS), Sydney, Australia, 11–13 December 2015; pp. 81–88.

24. Tomás, V.R.; Pla-Castells, M.; Martínez, J.J.; Martínez, J. Forecasting adverse weather situations in the road network. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2334–2343. [CrossRef]

25. Stamos, I.; Mitsakis, E.; Salanova, J.M.; Aifadopoulou, G. Impact assessment of extreme weather events on transport networks: A data-driven approach. *Transp. Res. Part D Transp. Environ.* **2015**, *34*, 168–178. [CrossRef]

26. Yu, R.; Abdel-Aty, M.A.; Ahmed, M.M.; Wang, X. Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 205–213. [CrossRef]

27. Tsirigotis, L.; Vlahogianni, E.I.; Karlaftis, M.G. Does information on weather affect the performance of short-term traffic forecasting models? *Int. J. Intell. Transp. Syst. Res.* **2012**, *10*, 1–10. [CrossRef]

28. Koesdwiady, A.; Soua, R.; Karray, F. Improving traffic flow prediction with weather information in connected cars: A deep learning approach. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9508–9517. [CrossRef]

29. Koetse, M.J.; Rietveld, P. The impact of climate change and weather on transport: An overview of empirical findings. *Transp. Res. Part D Transp. Environ.* **2009**, *14*, 205–221. [CrossRef]

30. Wang, X.; Zheng, X.; Zhang, Q.; Wang, T.; Shen, D. Crowdsourcing in its: The state of the work and the networking. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1596–1605. [CrossRef]

31. Chaniotakis, E.; Antoniou, C.; Pereira, F. Mapping social media for transportation studies. *IEEE Intell. Syst.* **2016**, *31*, 64–70. [CrossRef]

32. Xu, X.; Su, B.; Zhao, X.; Xu, Z.; Sheng, Q.Z. Effective traffic flow forecasting using taxi and weather data. In Proceedings of the ADMA 2016 Advanced Data Mining and Applications: 12th International Conference, Gold Coast, QLD, Australia, 12–15 December 2016; pp. 507–519.

33. Maghrebi, M.; Abbasi, A.; Rashidi, T.H.; Waller, S.T. Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, Spain, 15–18 September 2015; pp. 208–213.

34. Anantharam, P.; Barnaghi, P.; Thirunarayan, K.; Sheth, A. Extracting city traffic events from social streams. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 43. [CrossRef]

35. Ni, M.; He, Q.; Gao, J. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1623–1632. [CrossRef]

36. D'Andrea, E.; Ducange, P.; Lazzerini, B.; Marcelloni, F. Real-time detection of traffic from twitter stream analysis. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2269–2283. [CrossRef]

37. Zeng, K.; Liu, W.; Wang, X.; Chen, S. Traffic congestion and social media in China. *IEEE Intell. Syst.* **2013**, *28*, 72–77. [CrossRef]

38. Cao, J.; Zeng, K.; Wang, H.; Cheng, J.; Qiao, F.; Wen, D.; Gao, Y. Web-based traffic sentiment analysis: Methods and applications. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 844–853.

39. Tse, R.; Zhang, L.F.; Lei, P.; Pau, G. Social Network Based Crowd Sensing for Intelligent Transportation and Climate Applications. *Mob. Netw. Appl.* **2017**, *23*, 1–7. [CrossRef]

40. Haghighi, P.D.; Kang, Y.B.; Buchbinder, R.; Burstein, F.; Whittle, S. Investigating Subjective Experience and the Influence of Weather among Individuals with Fibromyalgia: A Content Analysis of Twitter. *JMIR Public Health Surveill.* **2017**, *3*, e4. [CrossRef] [PubMed]

41. Gaztelumendi, S.; Martija, M.; Principe, O.; Palacio, V. An overview of the use of Twitter in National Weather Services. *Adv. Sci. Res.* **2015**, *12*, 141–145.

42. Kirilenko, A.P.; Stepchenkova, S.O. Public microblogging on climate change: One year of twitter worldwide. *Glob. Environ. Chang.* **2014**, *26*, 171–182. [CrossRef]

43. Grasso, V.; Crisci, A. Codified hashtags for weather warning on twitter: An italian case study. *PLoS Curr.* **2016**, *8*. [CrossRef] [PubMed]

44. Park, K.; Lee, S.; Kim, E.; Park, M.; Park, J.; Cha, M. Mood and weather: Feeling the heat? In Proceedings of the 2013 ICWSM, Cambridge, MA, USA, 8–10 July 2013.

45. Li, J.; Wang, X.; Hovy, E. What a Nasty day: Exploring Mood-Weather Relationship from Twitter. In Proceedings of the 2014 ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 3–7 November 2014; pp. 1309–1318.

46. An, X.; Ganguly, A.R.; Fang, Y.; Scyphers, S.B.; Hunter, A.M.; Dy, J.G. Tracking climate change opinions from twitter data. In Proceedings of the Workshop on Data Science for Social Good, New York, NY, USA, 24 August 2014.

47. Cody, E.M.; Reagan, A.J.; Mitchell, L.; Dodds, P.S.; Danforth, C.M. Climate change sentiment on twitter: An unsolicited public opinion poll. *PLoS ONE* **2015**, *10*, e0136092. [CrossRef] [PubMed]

48. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 2013 International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.

49. Mikolov, T.; Yih, W.T.; Zweig, G. Linguistic regularities in continuous space word representations. In Proceedings of the 2013 HLT-NAACL, Atlanta, GA, USA, 9–14 June 2013.

50. Luo, Z. China's administrative region division reform and mechanism. *City Plan. Rev.* **2005**, *8*, 29–35.

51. Suh, B.; Hong, L.; Pirolli, P.; Chi, E.H. Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 20–22 August 2010; pp. 177–184.

52. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

53. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.

54. Smola, A.J.; Scholkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]

55. Kutner, M.H.; Nachtsheim, C.J.; Neter, J. *Applied Linear Regression Models*, 4th ed.; McGraw-Hill: Irwin, PA, USA, 2004; ISBN 0073014664.

56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

57. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. In Proceedings of the 2013 ECML PKDD Workshop: Languages for Data Mining and Machine Learning, Prague, Czech Republic, 23–27 September 2013; pp. 108–122.

58. Zhou, Z.H.; Feng, J. Deep Forest: Towards an Alternative to Deep Neural Networks. In Proceedings of the 2017 IJCAI, Melbourne, Australia, 19–25 August 2017.

59. Peng, Y.; Zhu, W.; Zhao, Y.; Xu, C.; Huang, Q.; Lu, H.; Zheng, Q.; Huang, T. Cross-media analysis and reasoning: Advances and directions. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 44–57. [CrossRef]

60. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake shakes twitter users: Real-time event detection by social sensors. In Proceedings of the 2010 International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 851–860.

61. Earle, P. Earthquake twitter. *Nat. Geosci.* **2010**, *3*, 221–222. [CrossRef]