The impact of Planned Special Events (PSEs) on urban traffic congestion

Ruwangi Fernando

Institute for Sustainable Industries & Liveable Cities, Victoria University, Melbourne, Australia

Abstract

The transport infrastructure of many cities has not been able to keep up with the pace of growth in the motorization rate or to counteract the intensification of urban traffic. Rising in traffic congestion in cities not only impacts the productivity costing billions but also responsible for more than 40% of all CO_2 emissions which results in global warming. While expansion and construction of new roads may be considered in some cases, in most, better management of existing infrastructure to lower traffic congestion is the only option.

The current state of the art commercial solutions can predict the recurring traffic situations, as this behavior can be easily learned from historical data. The challenge is to predict the non-recurrent congestion caused by events such as accidents, adverse weather, construction zones and Planned Special Event (PSE). Past research has shown, PSEs such as concerts or sports games, festivals and conventions has a huge impact on everyday urban transportation. Therefore, the aim of the proposed research is to investigate the impact PSEs on urban traffic congestion.

The proposed research is applying spatial-temporal big data mining methods to predict the impact of PSEs on the urban concession. Specifically, the proposed research will consider the characteristics of PSEs such as location, type, duration, audience and time and day of the event in the proposed analysis approach which enables to predict of urban congestion for future PSEs.

*Corresponding author. Email: Ruwangi.fernando@vu.edu.au

1. Introduction

Urban traffic congestion is an undeniable socio-economic problem that impacts every leading city in the world and indicates it will continue to worsen [1]. Urban congestion increases the travel time and pollutes the environment [2] which directly impacts the productivity, reliability, and well-being of the residents and visitors [3]. Urban traffic congestion caused by the recurring and non-recurring event [4]. The rush hour traffic is the recurring congestion and non-recurring congestion is congestion caused by planned events (e.g. special events, road construction) and unplanned events (e.g. Accidents, weather).

Cities organize Planned Special Events (PSEs) as these events promote their economy [5]. PSEs include cultural and sports activities, convention and exhibitions, entertainment events and commercial promotional activities [6]. One of the key elements of the success of these activities is transportation. PSEs are often held in short periods of the time, therefore, traffic impacts caused is temporary, so solving transportation problems caused by planned special activities will not require a permeant increase of road capacity [7] but definitely required to find better ways to mitigate congestion [8]. As PSEs are planned well in advanced and by predicting the congestion, we are able to mitigate the congestion better [9]. Which derives the main research question of the proposed study;

How to predict non-recurrent traffic congestion caused due to a PSE?

The last few years have been marked as the beginning of exploring the Big data in all industries. This had an enormous impact in the transport sector where it contributes towards the development of ITS (Intelligent Transportation System). Previous researchers have used data mining techniques on taxi trajectory data to measure parameters such as the travel time [10], the average speed of the road section [11], the congestion degree [12], the best route selection[13], measures to seek the next passengers [14], provide real-time urban traffic flow distribution [15] are some other topics used big data mining approaches. These research were possible due to evolved transportation data collection methods such as access to real-time traffic information using the static sensors and GPS (Global Positioning System) devisers which capture the evolution of urban activity in real time [10]. The static sensors are installed directly in the roadway (e.g. magnetic loops, cameras) to collect data which are necessary but also inherent limitations such as partial coverage of a road network and high installation and maintenance costs. Alternatively, GPS data also referred to as Floating Car Data received from GPS embedded vehicles or smartphones is considered as the most suitable solution to overcome the limitations of fixed sensors [15]. The significant growth of GPS data collection methods has allowed many useful research and practical insights on urban traffic congestions such as automatic detection of road events in real time [16], evaluation of the duration future congestions and the prediction of road events [3, 17].

With the rapid development of taxi industry, taxis have made up for the shortcomings of public transport and private transport [18]. It has become one of the essential means of transport for people's daily travel, and have played an important role in urban travel especially for nonroutine travels such as getting to and from a planned special event [7]. With the introduction of GPS technology taxi trajectory data became the lifeline of intelligent transportation and urban traffic management [19]. Taxi GPS data records the dynamic changes of urban traffic and crowd movement, which provides a large amount of relevant data to research on urban traffic [13].

The proposed research aim is to predict urban non-recurrent traffic congestion influenced by planned special events using big data mining techniques. The detailed research objectives are listed below:

- Measure urban non-recurrent traffic congestion influenced by PSEs.
- Investigate the relationship between non-recurrent traffic congestion and characteristics of PSEs (the explanatory variables or covariates)
- Develop a decision-making tool that can be used to predict urban traffic congestion influenced by PSEs.

The proposed research uses Sydney, Australia as a case study to develop and evaluate the methodology. Traffic congestion in Sydney is currently among the highest in the world along with New York, Toronto, and Philadelphia. The annual cost of Sydney's congestion is $6.1 billion in 2015, making Sydney the most congested city in Australia and its estimated to increase up to $12.6 billion by 2030. This value includes the productivity lost by those unable to access employment due to increased travel times, the cost to business and their customers and the environment impact of congestion.

2. Contribution to knowledge

Scholars have been at the forefront of recurring congestion research and have published a large number of papers with academic impact [20]. However, there is a lack of research on non-recurring traffic congestion and especially on PSEs. The limited published research on PSEs influence on traffic congestion cover mega-events such as Pope's visit [9], large concerts, Olympic games [8] and big soccer matches [5]. Therefore, the proposed research will fill this gap by analysing the influence of different categories of urban PSEs on urban traffic congestion.

PSEs characteristics such as type, location, time of occurrence, duration, no of expected audience, market area and audience accommodation define its variability [21]. One gap in the existing literature is to analyse these characteristics of PSEs to find its influence on urban traffic congestion [22]. Therefore, the proposed research major contribution is to fill this gap by developing a model to analysis PSEs characteristics and measure its influence on urban traffic congestion.

Proposed research use data from Sydney and therefore this becomes the pioneer study in Australia to use data mining methods to analyse the characteristics of PSEs to measure the impact of PSEs on urban traffic congestion.

3. Statement of significance

Surveys of the inhabitants of large cities reveal a universal dissatisfaction with urban traffic congestion. Governments spend a lot trying to monitor and understand urban congestion but failed due to its complexities and unpredictability. If developing new infrastructure is not an option, better management of urban traffic congestion is the solution for urban congestion. Traffic flow in the urban road networks is very sensitive to both recurring and non-recurring events. Kwoczek, et al. [5] propose a vision to predict future congestion by providing a holistic view of urban congestion. The vision requires capturing all the influencing factors on traffic and their overlapping situations. Proposed research will contribute to the development of this vision by analysing the influence of PSEs on urban traffic congestion.

This pioneering study will demonstrate how the characteristics of PSE influence urban traffic congestion.

This analysis will enable the prediction of future urban traffic congestion influenced by PSEs. Therefore, urban authorities can use these predictions to implement temporary travel solutions by providing dedicated lanes / special lanes to ease the congestion caused by PSEs. Hopefully, more research will follow and improve forecasting ability significantly.

PSEs temporarily cause two subsequent waves of congestion around the venues [22]. First, due to people arriving at the venue and the second due to people leaving from the venue. Previous researchers have predicted and visualized the impact of the second wave of traffic by analysing the first wave using existing congestion analysis methods and with the assumption of outbound congestion should be similar to inbound congestion caused by PSEs. It is more important to predicting both the inbound and outbound congestion in advance as it will provide lead time for transportation authorities to plan and implement a successful congestion mitigation strategy. Therefore, proposed research becomes pioneer research in contributing a novel methodological approach to analysis and predict both inbound and outbound congestion caused by PSEs.

Most previous research on urban congestion use freeway congestion data which does not exactly reflect the urban traffic congestion especially when analysing the influence of PSEs. Compared to the freeway the inner city consists of thousands of roads and intersections and therefore proposed research will involve a detailed analysis of the road structure of the of the city of Sydney, Australia.

4. Literature Review

The proposed research is a cross-disciplinary study of urban traffic congestion, planned special events, spatial-temporal big data mining methods and congestion data visualisation approaches. This section will introduce the development of the literature in all the above three areas.

4.1 Urban Traffic Congestion

Traffic congestion is a state of the network when traffic demands exceed the available capacity. The state when traffic demand equals capacity is known as 'saturation'. This state results in lengthy delays and queue formation until demands reduce to levels below capacity [23]. The capacity of a network is not static, but variable and depends on many factors, including traffic volumes and flow conditions on each component of the network, road link conditions, traffic signal phasing and cycle times, parking activity, and other various factors [24]. During periods of traffic congestion, small disruptions to traffic flow can result in dramatic reductions in vehicle speeds with stop/start conditions propagating back into the traffic flow [23].

Traffic congestion within an element of the network is simplistically described in the speed flow diagram in Figure 1. As the network demand flow increases, vehicle speeds reduce until they reach a maximum volume (shown as qmax in Figure 1) [23]. As traffic demands then increase beyond this point, vehicle speeds reduce further, thus causing reduced flow [24]. This results in unstable queue formation and lengthy delays within the network. It is also very important to understand that due to the non-learner nature of this relationship of volume and speed if the congestion can be well managed by reducing the volume somewhat. Further congestion is distinguished between two forms of this phenomenon, recurrent congestion and non-recurring congestion [25].
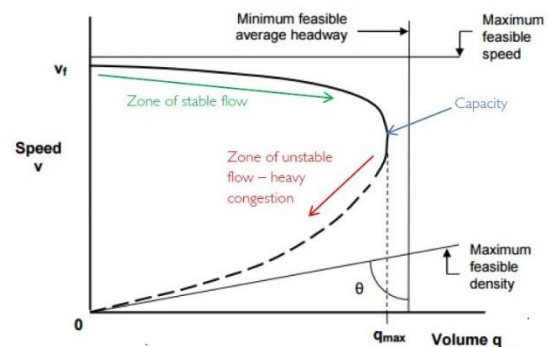

Figure 1. Speed flow diagram
Source: Austroads, Guide to Traffic Management, Part 2: Traffic Theory

Recurrent congestion occurs in time intervals associated with peak hours where the demand for traffic is very close or even exceeds the available road capacity. Therefore, recurrent congestion tends to be predictable and repetitive. Many studies have proven its predictable nature using time series analysis.

According to Bennett [23] **Non-recurring congestion** results from random events or with difficulty predictable that vary from one road segment to another. The main events that are at the origin of the non-recurring congestion are incidents in traffic (accidents, breakdowns, etc.), weather conditions, road maintenance and PSEs. Non-recurring congestion is sometimes associated with the reduction in road capacity due to accidents, road maintenance but also, its associated with high demand due to Planned Special Events.

Harmful effects of congestion

Economic consequences - Road congestion causes extra time to move from point of origin to point of destination. This additional time influences the delivery times of goods as well as travel time. In addition, traffic congestion reduces the labour hours as well as accessibility to economic activities (Robitaille and Nguyen, 2003). In Sydney, congestion is $6.1 billion in 2015, and its estimated to increase up to $12.6 billion by 2030.

Ecological consequences -When the number of vehicles (density) increases on the road, the speed decreases, and the time of displacement is prolonged. This results in an additional emission of air pollution as well as noise pollution [26]. The residents of the road and the motorists are most affected by these emissions. In this context, several studies have been conducted, by Hu, et al. [27] who showed that air quality deteriorates due to traffic.

Social consequences -In addition to the ecological and economic consequences, congestion has a detrimental effect on society. Speed reduction can lead to a decrease in social contacts between people, especially when the tolerated travel time is exceeded because of congestion [27]. In addition to the ecological consequences, atmospheric pollution additional increase due to this phenomenon has negative impacts on health, for example the effect of some gases on the respiratory capacity. In addition, noise pollution affects the physical state of people (stress, quality of sleep, etc.) [26, 27].

Due to the above discussed many consequences of traffic congestion, it derives into several definitions of indicators or ways to measure traffic congestion. Many of these traffic congestion indicators have been reported in the literature. The [28] identified these indicators among its member countries and categorized them into six major categories indicators:

- Speed-based indicators
- Capacity indicators, service level
- Temporal indicators based on the delay
- Spatial indicators
- Reliability indicators
- Economic cost/efficiency indicators

The purpose of an indicator is to assess a phenomenon using its measurable parameters. Proposed research uses the temporal indicators based on the travel time or delay as it is the most used congestion indicator by transport researchers and engineers [29].

The calculation of the delay generally refers to the difference between the current travel time and the free circulation time on a section of the road or between two locations/zones. This Can be determined using different methods adapted to a wide variety of situations, such as a freeway or arterial system.

The delay may be "recurring" or "Non-recurring". The Recurring delays are encountered daily during rush hour travel and which can be inferred from historical data. Non-recurring delays are those caused by events, incidents or accidents and are generally divided into two periods, the immediate time and the residual time for events such as accidents and pre and post congestion time due to PSPEs. The former refers to the delay in the incident while the second is the delay after the incident was processed and the roads were restored.

4.2 Planned special events

In recent years, government departments and non-governmental organizations organise PSEs to promote the image, developed tourism, promote distinctive cultures, extend to continuing traditional values, driving industrial and commercial interests etc. Planned special events (PSE) include sporting events, concerts, festivals, and conventions occurring at permanent multi-use venues as well as less frequent public events occurring at temporary venues, such as parades, fireworks display, bicycle races, sporting games, motorcycle rallies, and seasonal festivals. The growing number of PSEs can stimulate industrial and commercial interests, but cause severe road traffic congestion, unbalanced parking supply and demand, serious parking violations, severely intertwined moving lines, impeding smooth traffic and safety [5].

PSEs usually happen in dense inner-city areas with complex traffic networks. Most previous research on urban congestion use freeway congestion data with does not exactly reflect the urban traffic congestion especially when analysing the influence of PSEs [5, 9].

Unlike other causes of non-recurrent congestion, PSEs are unique due to its many characteristics. Therefore, the PSE characteristics can be used to classify the PSEs into categories such as a discrete/recurring event at a permanent venue, continuous event, street use event, regional/multi-venue event and as a rural event and many other as illustrated by figure 2 [30].
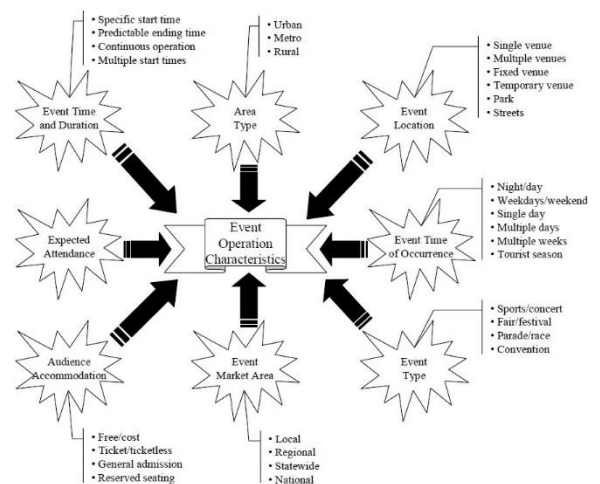

Figure 2. Planned Special Events operation characteristics

Source: Event studies: Theory, research, and policy for planned events [21]

To manage the intense travel demands of PSEs and to maintain transportation systems safety, mobility, and reliability local government and host organization must work together and establishes a special congestion mitigation plan to accommodate transportation demand. Then the categories impact on urban congestion can be used to predict the congestion of future PSEs.

As PSEs are planned well in advance and therefore by analysing historical data of these PSEs if the congestion can be predicted for future PSEs, it will inform the transportation authority's appropriate cause of action. It will minimize the inconvenience to the residents and improve the accessibility of the event attendees.

In Sydney Australia, the number of large-scale events is increasing, and in order to hold successfully, it is necessary to improve the convenience of use as well as the development of contents. In order to minimize the inconvenience of the residents and the convenience of visitors, it is necessary to establish transportation measures to arrive at the PSEs and departure after the PSEs.

4.3 Spatial-temporal big data mining

The spatial and temporal relationships in spatiotemporal data are usually complex. Spatial data refers to the data used to represent the geographical location. Temporal (time) data refers to the data related to the time series and expresses the changes of the target events over time.

The term Spatial-temporal big data is not clearly defined scientifically or technically. Following the definition of Big data, Spatial-temporal big data also should possess one or more of the following characteristics:

- Volume (amount of data)
- Velocity (speed of data creation)
- Variety (heterogeneity of data)
- Veracity (reliability of data, eg when using social media content)

Data Mining is used with the goal to find pattern large databases and recognize structures that were previously unknown [31]. This way, data Mining well from the subject of machine learning, in which mainly known patterns should be recognized in new records. The term data mining can be traced back to the 1950s when once large databases were searched using machine learning techniques. Over time, techniques evolved to be associated with data mining let become an integral part of modern database technologies and play an important role in the transition from classical data management Software tools that can directly support management decisions [32].

The objectives that are pursued in data mining projects are largely the same across different context. It is to provide descriptive, diagnostic, predictive and prescriptive

analysis. Also, they follow a similar data mining process typically consists of a sequence of individual process steps:

- Data integration and clean-up
- Data selection and transformation
- Search patterns and trends
- Knowledge finding and construction

Due to its the interactive and iterative nature of its process, it's not easy to define the limits of data mining. As data mining does not necessarily end with the analysis of data using algorithms but also includes visualizations, which enhancers the observers understand of the new knowledge.

In recent years, spatial-temporal data mining has become a research hotspot in the field of data mining [33] and has won widespread attention in many fields, such as traffic management, crime analysis, disease surveillance, environmental monitoring and public health. Spatial-temporal data, as the name implies, necessarily includes data related to the geographical location of the space and to time series.

Existing knowledge of trajectory mining

Mining trajectory data can directly reflect the urban congestion status, traffic time estimation, and traffic anomaly detection. Intelligent transportation is the most direct application of trajectory data. Bi-Yu, et al. [34] used taxi trajectory data for short-term traffic forecasting; Wang, et al. [35] used the data to the estimation of travel time which resulted in higher time accuracy. Some unusual behaviours in the traffic flow through trajectory detection also achieved good experimental results. Pang, et al. [36] conducted a study on the rapid detection of traffic anomalies using the LRT model. Shanthi and Ramani [37] used trajectory data to study traffic congestion monitoring and traffic anomalies. Mao, et al. [38] probed the spatiotemporal causation of traffic anomalies in traffic flow; Chawla, et al. [39] deduced the root cause of traffic anomalies; Pan, et al. [40] conducted congestion perception in traffic anomalies based on crowd movement and social networks Aspects of research. At present, the speed of uploading and processing trajectory data is in demand to meet the real-time requirements. The time estimation and anomaly detection of the above research are based on historical data.

Trajectory data can also be used to optimize and design traffic routes. Taxis, as an important tool for the movement of urban populations, taxi drivers are often familiar with road conditions and their driving directions can be considered as the optimal route (classic route) between two points. Yuan, et al. [41] calculated the fastest and most personalized route design based on the taxi track data in the T-Drive project, saving not only 5 minutes of driving time every 30 minutes but also different roads for different users Relieve possible congestion. At the same time, through the optimal route comparison can also find unreasonable road planning, provide the basis for the design of urban traffic

routes [4]. Using taxi GPS trajectory data, Castro, et al. [15] proposed a method to construct a taxi density model, which can predict future traffic conditions and estimate the impact of vehicle emissions on urban air quality. Chen, et al. [42] proposed the use of taxi night GPS track data to plan night bus routes. For bi-directional bus, they proposed two-stage route selection. The hot spots in the area were detected by detecting passengers getting on and off.

Existing knowledge of GPS trajectory data mining approaches

It is possible to perform various semantic analysis by using mining techniques on GSP data. The initial big data mining study was to develop a new mechanism to extract relevant information from large datasets. However, as more and more large types of data, especially those spatial-temporal relationships, requires complex calculations. Previous researchers have use association analysis, cluster analysis, and classification analysis as spatial-temporal data mining approach mostly on the GSP analysis. Table 1 summarizes the GSP data mining literature by the data mining approach.

Table 1: Summary of literature on data mining algorithm used on GPS data

Data Mining Approach	Algorithm	Reference
Association	Apriori algorithm	[43], [44], [45]
	Bayesian networks	[46], [47], [48]
	Binary decision tree	[12]
	Association rules	[49], [12], [50]
Clustering	K-means	[14], [51]
	DBSCAN clustering algorithm	[52], [11]
Classification	SVM (Support Vector Machine)	[2], [53], [54] [55]
	Neural Networks (NN)	[56], [57]
	Decision trees	[58]
	Naive Bayes	[59]
	Random Forests (RF)	[60]

Association analysis is a method of finding an association rule hidden in a large amount of data [61]. The results of association analysis are not explicitly expressed but are meaningful patterns or rules that can provide a good clue for future analysis.

Apriori algorithm is an algorithm for associative rule learning and a set of items frequently occurring in transactional databases [43]. This proceeds by identifying frequent individual items in the database and extending them to a larger set of items as the itemset appear sufficiently frequently in the dataset. The Apriori determined the frequency of the general trend of item set databases, which can be used to determine association rules. Liu and Huang [45] used the apriori algorithm to analyse the spatial relationship between the spatial pattern of the traffic accident and neighbouring space. In this study, the objects within 500m from the traffic accident point are considered as the space object adjacent to the traffic accident, and the road facilities, cultural facilities, industrial facilities, service facilities, medical welfare Facilities, housing facilities, administrative institutions, topography, and rivers. Another representative algorithm in association analysis is Bayesian networks.

The Bayesian network is a graphical model of the probability of expressing the set as conditional independence through a set of random variables and a directed acyclic graph [48]. In Bayesian networks, there are efficient algorithms to perform reasoning and learning [47]. The generalization of the Bayesian network [62], which can express problems and solve the problem under uncertainty, is called an impact diagram. Xiao, et al. [46] analysed Bayes network analysis from smart mobile and analysed the personal presence of transport mode selection (walk, bike, e-bike, bus, and car).

Clustering analysis uses iterative techniques to group cases of datasets into clusters with similar characteristics. This grouping is useful for exploring data, identifying incorrect parts of data, and creating predictions. The transport discipline Pan, et al. [52] finds patterns from taxi trajectory data, classifies users, and clusters them into groups with similar views.

Analysis based on spatial variables users k-means [14, 51], nearest neighbour, kernel density analysis, DBSCAN algorithm [11] [52]. Liu, et al. [63] grouped user location information using a grid-based clustering algorithm using user logs and location information collected from mobile. stay points, and finally predict or recommend the user's point of interest (POI).

Niu, et al. [64] used GIS techniques on traffic accident data and spatial patterns of traffic accidents and spatial relationships between neighbouring spaces were studied. Spatial patterns were analysed using the k-means algorithm. The k-means clustering is a distance-based clustering technique that groups data near a reference point into a cluster. The k-means clustering divides data into k clusters so that each point belongs to the closest cluster. Therefore, this method is useful for classifying data or learning new data from existing data. The traffic accidents are clustered by individual attributes (accident content, type of accident, the day of occurrence, month, day and night, accident type, road type, road width, drunk driving, land use, weather) and analysed the characteristics of each cluster.

Lu, et al. [11] performed cluster analysis using density analysis and kernel analysis method using traffic accident data from 1st January to 31st December 2013, Industrial complexes, and residential areas and analysed that the traffic accidents of the three cars were larger than those of the undercarriages. Pan, et al. [52] reported that DBSCAN algorithm can be successfully used for clustering the characteristics of user behaviours.

Classification analysis has been applied in various fields. Recently, it has also been used to recognize emotions and stresses and to classify social relations and personality. Most used classification approaches are SVM (Support Vector Machines) [53] , RF (Random Forests) [60], NN (Neural Networks) [56], Naive Bayes and decision trees algorithms [59].

Kong, et al. [2] visualized the data collected from the subway transportation card to intuitively confirm the density of the floating population in the subway transportation network by time and place. Kumar, et al. [55] analyses the spatiotemporal data collected from the mobile through the sequential pattern mining to find the life patterns of the users.

Tang, et al. [59] used a Bayesian method is used for prediction of routes and preferences of transport of travellers with algorithms that have the power to learn using a dynamic knowledge base fuelled by user behaviour in a certain period of time. The project carried out by Xiao, et al. [46] implements a system on a Bayesian network about K2 learning algorithms that allow extracting knowledge information prior and use advanced mining techniques of data. The system extracts the information from GPS data and through similarity techniques are estimate modes of transport. The network identifies with great accuracy the rate changes of speed and trajectories on routes, being able to adjust the GPS parameters in factors of similarity for the adequate detection of transport mode. In He, et al. [65] performs a review on the formal methods for the recollection of GPS data and detection of modes of transport proposing a methodology of Acquisition, filtering and data processing a through android platforms and the possible algorithms for detecting modes of transport. His proposal offers a panorama of system from the client's perspective and other of the server and the system is based on processes in real time and how they could be optimized algorithms and energy consumption in the smartphones. Also, Feng and Timmermans [66] presents a comparison about algorithms used to extract acceleration patterns that identify modes of transport over segments with length and time limits. His study shows a taxonomy of algorithms based on Bayesian networks, regressions linear, vector machine and decision tables.

De Brébisson, et al. [56] develop a system of inference of modes of transport with algorithms in neural networks based in GPS data from Microsoft's Geolife project Asia (Microsoft Research, 2012). This method is to customize the speed profiles of each mode of transport in such a way

that the same system learns if there is a variation of speed and can classify each activity of travel according to the best means of transport that fits. The main contribution of this research is a framework for the construction of user mobility speed profiles with scattered and unfiltered GPS data based on velocity characteristics restrictions of a transport network.

4.4. Visualization of congestion data

Visualization can directly relate users to data by enabling users to interact with data in a simple visual manner [67]. Data visualization greatly improve the efficiency and accuracy of analysis and decision making as it blends user intelligence and machine intelligence together [68]. Taxi trajectory data is one of the most common traffic data, each track record contains not only location information but also recorded the time. The proposed research aims to visualize taxi trajectory data which provides more semantic information about urban traffic routes to enhances our understanding of anomalous trajectories generated by planned special events. Traffic data visualization methods are mainly divided into three categories: statistical heat map, space-time trajectory and multi-dimensional coding [69].

Statistical heat map -Is one of the most basic and common visual forms, it is usually used to express the distribution of a single value (such as traffic flow, number of people, etc.) in different locations [70]. The colour depth of each pixel on the map represents the average hourly operation of the location; the deeper the colour, it represents a frequency of the taxi boarding for the location. Therefore, from the colour distribution of the thermograms the intuitively provides information which urban areas are busier.

Space-time trajectory -Trajectory data contains a wealth of space-time information, we can visualize the trajectory data from the perspective of time and space [69]. Trajectory time attributes are mainly linear time and periodic time two kinds. Linear time can be coded using a visualization method based on a timeline that encodes the start of the data at both ends. The view uses a timeline approach to visualize the relationship between subway route selection and duration. Starting from one station, users can choose any one station to get off according to the subway network, and the length on the horizontal axis represents the time spent on the whole trip. For periodic times, such as weeks, days, and hours, the most common visualization method is a ring layout.

Multidimensional coding - Visual forms such as thermograms and trajectories generally only encode less dimensional information. As the data dimension becomes larger, the common visual form becomes more difficult to navigate with such complex information and therefore design appropriate visual coding for application scenarios

and analytic task pertinence. For example, the Space-Time Cube (STC) is a commonly used method of expressing spatiotemporal trajectories, where the trajectory of an object is expressed using lines that extend gradually upward from the map plane [71]. To show various attributes (such as crowd type, vehicle type, event occurrence details, etc.) in different positions of the track, colour, dots, geometric shapes or specially designed icons can also be added on the track lines [72].The polyline encodes the vehicle's position in space and time dimensions, where the colour from red to green encodes the speed of movement and some traces of traffic jams are easily discernible.

The visual analysis technology used in traffic intelligence analysis system can be further divided into three categories according to the type of application: query, statistical analysis and query reasoning.

5. Proposed Methodology

This section describes the methodological approach to carry out for this proposed project and the data used.

Proposed reseach use the duration of traffic congestion or delay caused by PSE's as a quantitative indicator to objectively reflect the impact of PSE's on traffic operations. Reliable traffic duration prediction can provide effective guidance and information for traffic control systems, traffic guidance systems, and traveler service systems. Which enable the relevant management departments to take necessary traffic management control measures such as induce the driver to choose the driving routes or to make public transpotation effectively avaliable. Therefore, proposed research methodology develops a prediction model using congestion data, Travel Zones (TZs) data combined with data about PSEs.
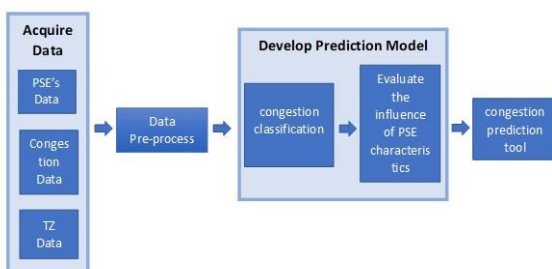

Figure 3. Proposed methodology

The travel time congestion data is extracted from Uber movement and from few web-based mapping services such as Google, Bing, Here and Open street maps using their respective distance matrix APIs. Along with Travel Zones (TZs) data which provide spatial information of the study area of Sydney is collected from open data portal of the transport of NSW website combined with PSEs data extracted from the Live Traffic Hazards API provided by Live Traffic NSW web site using the developer APIs. Previous researchers have adopted survival analysis when duration prediction models are developed in many research domains, starting with medical research to study of the time that has elapsed since the diagnosis of a disease until the cure or death, economic research to study of the time spent between the end of studies and the beginning of a first job, engineering research to study of the time between the commissioning of an electrical component and its failure, in social science research to study of the duration before the first marriage and transport research to study of the clearance time after a road accident. Survival analysis consists of a taxonomy of survival analysis methods to analyse the distribution of time elapsed until the occurrence of a specific event [74]. It is also referred to as Hazard Duration or Time-to-event or Time to death. Therefore, proposed research applies random survival forests in predicting the traffic duration influenced by PSEs traffic duration prediction. Ishwaran, et al. [75] proposed the random survival forest. It is an extension of random forests in survival analysis. It is mainly applied to censored survival data. Survival random forests not only has the advantage of random forest and survival analysis but also overcome the limitations of traditional survival analysis as they rely on mandatory assumptions.

6. Conclusion

It is a huge challenge to make urban mobility more accessible, sustainable, efficient and environmentally friendly. The harmful effects of congestion fall directly on all the users of the transport system. The non-recurring congestion occurs irregularly which makes it difficult to predict and specially PSEs hasn't had much research attention but has been impacting urban transportation heavily. Therefore, proposed research uses the latest available technology, data and research methods to predict the traffic congestion influenced by PSEs. Applying suitable data mining algorithms, the congestion indicators are calculated for the inbound and outbound traffic generated by PSEs. Theses congestion indicators are then analysed with the PSEs characteristics to gain insightful information which can be used to predict traffic delays for future PSEs. The outcome of the proposed research will help city traffic authorise in mitigating congestion influence by PSEs also government bodies can use this information in their strategic decision-making process of urban planning and development.

References

[1]   J. Zhao, R. Li, X. Liang, and K. Xu, "Second ring is the most jammed road in Beijing: A view from taxis and Weibo," in Service Systems and Service Management

(ICSSSM), 2016 13th International Conference on, 2016, pp. 1-6.

[2]   X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," Future Generation Computer Systems, vol. 61, pp. 97-107, 2016.

[3]   Q. Yang, Z. Gao, X. Kong, A. Rahim, J. Wang, and F. Xia, "Taxi Operation Optimization Based on Big Traffic Data," pp. 127-134, 2015.

[4]   Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in Proceedings of the 13th international conference on Ubiquitous computing, 2011, pp. 89-98.

[5]   S. Kwoczek, S. Di Martino, and W. Nejdl, "Stuck around the stadium? an approach to identify road segments affected by planned special events," in Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, 2015, pp. 1255-1260.

[6]   T. Absalyamov, "The Influence of Cultural and Sport Mega-events on Sustainable Development of the City," Procedia - Social and Behavioral Sciences, vol. 188, pp. 197-201, 2015.

[7]   F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, "Using Data From the Web to Predict Public Transport Arrivals Under Special Events Scenarios," Journal of Intelligent Transportation Systems, vol. 19, pp. 273-288, 2014.

[8]   R. Li and M. Guo, "Effects of odd–even traffic restriction on travel speed and traffic volume: Evidence from Beijing Olympic Games," Journal of traffic and transportation engineering (English edition), vol. 3, pp. 71-81, 2016.

[9]   S. Sharma, L. D. Galicia, J. Susen, and L. E. Cornejo, "Employing big data analytics to monitor transportation impacts in El Paso during Pope's visit to the US-Mexico border City of Juárez," International Journal of Transportation Science and Technology, vol. 6, pp. 127-142, 2017.

[10]   I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, et al., "The role of big data in smart city," International Journal of Information Management, vol. 36, pp. 748-758, 2016.

[11]   Y. Lu, S. Xiang, and W. Wu, "Taxi Queue, Passenger Queue or No Queue?," in Proc. of 18th International Conference on Extending Database Technology (EDBT). Brussels, Belgium, 2015, pp. 593-604.

[12]   Y. Araya, "Detecting Swiching Points and Mode of Transport from GPS Tracks," ed, 2012.

[13]   X. Luo, L. Dong, Y. Dou, N. Zhang, J. Ren, Y. Li, et al., "Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China," Journal of Cleaner Production, vol. 142, pp. 926-935, 2017.

[14]   X. Liu, J. Biagioni, J. Eriksson, Y. Wang, G. Forman, and Y. Zhu, "Mining large-scale, sparse GPS traces for map inference: comparison of approaches," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, pp. 669-677.

[15]   P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in International Conference on Pervasive Computing, 2012, pp. 57-72.

[16]   T. Kim, B. Park, J. Park, and H.-G. Cho, "An Efficient Clustering Algorithm for Massive GPS Trajectory Data," Journal of KIISE, vol. 43, pp. 40-46, 2016.

[17]   N. Lin, "Path Planning Method Based on Taxi Trajectory Data," Journal of Information and Computational Science, vol. 12, pp. 3395-3403, 2015.

[18]   Y. Hu, Y. Yang, and B. Huang, "A Comprehensive Survey of Recommendation System Based on Taxi GPS Trajectory," pp. 99-105, 2015.

[19]   K. Lai and Y. Li, "Image Impacts of Planned Special Events: Literature Review and Research Agenda," Event Management, vol. 18, pp. 111-126, 2014.

[20]   S. Kwoczek, S. Di Martino, and W. Nejdl, "Predicting Traffic Congestion in Presence of Planned Special Events," in DMS, 2014, pp. 357-364.

[21]   D. Getz and S. J. Page, Event studies: Theory, research and policy for planned events: Routledge, 2016.

[22]   D. Leilei, S. Zheng-liang, G. Jin-gang, and Q. Hong-tong, "Study on traffic organization and management strategies for large special events," in System Science and Engineering (ICSSE), 2012 International Conference on, 2012, pp. 432-436.

[23]   D. Bennett, Guide to traffic management: part 2: traffic theory, 2008.

[24]   C. Han, J. Luk, and K. Bevrani, Guide to traffic management part 2: traffic theory, 2015.

[25]   M. Lu, C. Lai, T. Ye, J. Liang, and X. Yuan, "Visual Analysis of Multiple Route Choices Based on General GPS Trajectories," IEEE Transactions on Big Data, vol. 3, pp. 234-247, 2017.

[26]   M. Wardman and A. L. Bristow, "Traffic related noise and air quality valuations: evidence from stated preference residential choice models," Transportation Research Part D: Transport and Environment, vol. 9, pp. 1-27, 2004.

[27]   D. Hu, J. Wu, K. Tian, L. Liao, M. Xu, and Y. Du, "Urban air quality, meteorology and traffic linkages: Evidence from a sixteen-day particulate matter pollution event in December 2015, Beijing," Journal of Environmental Sciences, vol. 59, pp. 30-38, 2017.

[28]   European Conference of Ministers Transport and Research Centre, Managing urban traffic congestion: Organization for Economic, 2007.

[29]   C. Liu, K. Qin, and C. Kang, "Exploring time-dependent traffic congestion patterns from taxi trajectory data," in 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2015, pp. 39-44.

[30]   L. K. Jago and R. N. Shaw, "Consumer perceptions of special events: a multi-stimulus validation," Journal of Travel & Tourism Marketing, vol. 8, pp. 1-24, 2000.

[31]   M. Peng, D. Chen, Q. Xie, Y. Zhang, H. Wang, G. Hu, et al., "Topic-net conversation model," in International Conference on Web Information Systems Engineering, 2018, pp. 483-496.

[32]   M. Peng, Q. Xie, Y. Zhang, H. Wang, X. Zhang, J. Huang, et al., "Neural sparse topical coding," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 2332-2340.

[33]   J. Huang, M. Peng, H. Wang, J. Cao, W. Gao, and X. Zhang, "A probabilistic method for emerging topic tracking in microblog stream," World Wide Web, vol. 20, pp. 325-350, 2017.

[34]   C. Bi-Yu, W. H. Lam, Q. Li, A. Sumalee, and K. Yan, "Shortest path finding problem in stochastic time-dependent road networks with stochastic first-in-first-out property," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 1907-1917, 2013.

[35]   Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 25-34.

[36]   L. X. Pang, S. Chawla, W. Liu, and Y. Zheng, "On mining anomalous patterns in road traffic streams," in International Conference on Advanced Data Mining and Applications, 2011, pp. 237-251.

[37]   S. Shanthi and R. G. Ramani, "Classification of vehicle collision patterns in road accidents using data mining algorithms," International Journal of Computer Applications, vol. 35, pp. 30-37, 2011.

[38]   F. Mao, M. Ji, and T. Liu, "Mining spatiotemporal patterns of urban dwellers from taxi trajectory data," Frontiers of Earth Science, vol. 10, pp. 205-221, 2015.

[39]   S. Chawla, Y. Zheng, and J. Hu, "Inferring the root cause in road traffic anomalies," in Data Mining (ICDM), 2012 IEEE 12th International Conference on, 2012, pp. 141-150.

[40]   B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013, pp. 344-353.

[41]   J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, et al., "T-drive: driving directions based on taxi trajectories," in Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems, 2010, pp. 99-108.

[42]   C. Chen, D. Zhang, N. Li, and Z.-H. Zhou, "B-Planner: Planning bidirectional night bus routes using large-scale taxi GPS traces," IEEE Transactions on Intelligent Transportation Systems, vol. 15, pp. 1451-1465, 2014.

[43]   L. Wang, K. Hu, T. Ku, and J. Wu, "Frequent spatiotemporal trajectory pattern mining based on pheromone concentration," Journal of Information & Computational Science. vol. 10, pp. 645-658, 2013.

[44]   Y. Zhao, Y. Liu, X. Yu, D. Wei, C. Shan, and Y. Zhao, "Method for mobile path prediction based on pattern mining and matching," Journal of Jilin University (Engineering and Technology Edition), vol. 38, pp. 1125-1130, 2008.

[45]   Z. Liu and Y. Huang, "Mining co-locations under uncertainty," in International Symposium on Spatial and Temporal Databases, 2013, pp. 429-446.

[46]   G. Xiao, Z. Juan, and C. Zhang, "Travel mode detection based on GPS track data and Bayesian networks," Computers, Environment and Urban Systems, vol. 54, pp. 14-22, 2015.

[47]   T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, 2010, pp. 1625-1631.

[48]   J. Forbes, T. Huang, K. Kanazawa, and S. Russell, "The batmobile: Towards a bayesian automated taxi," in IJCAI, 1995, pp. 1878-1885.

[49]   D. Xia, X. Lu, H. Li, W. Wang, Y. Li, and Z. Zhang, "A MapReduce-Based Parallel Frequent Pattern Growth Algorithm for Spatiotemporal Association Analysis of Mobile Trajectory Big Data," Complexity, vol. 2018, 2018.

[50]   P. Gayathiri and B. Poorna, "Effective Gene Patterned Association Rule Hiding Algorithm for Privacy Preserving Data Mining on Transactional Database," Cybernetics and Information Technologies, vol. 17, pp. 92-108, 2017.

[51]   M. Lin and W.-J. Hsu, "Mining GPS data for mobility patterns: A survey," Pervasive and Mobile Computing, vol. 12, pp. 1-16, 2014.

[52]   G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi GPS traces," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 113-123, 2013.

[53]   J. G. Lee, J. Han, X. Li, and H. Cheng, "Mining discriminative patterns for classifying trajectories on road networks," IEEE Transactions on Knowledge and Data Engineering, vol. 23, pp. 713-726, 2011.

[54]   H. Tang, "Urban Transportation Analysis Using Taxi Trajectory and Weather Data," Kent State University, 2016.

[55]   D. Kumar, H. Wu, Y. Lu, S. Krishnaswamy, and M. Palaniswami, "Understanding Urban Mobility via Taxi Trip Clustering," in 2016 17th IEEE International Conference on Mobile Data Management (MDM), 2016, pp. 318-324.

[56]   A. De Brébisson, É. Simon, A. Auvolat, P. Vincent, and Y. Bengio, "Artificial neural networks applied to taxi destination prediction," arXiv preprint arXiv:1508.00021, 2015.

[57]   C. Chen, D. Zhang, P. S. Castro, N. Li, L. Sun, S. Li, et al., "iBOAT: Isolation-based online anomalous trajectory detection," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 806-818, 2013.

[58] Y. Zhu, Y. Zheng, L. Zhang, D. Santani, X. Xie, and Q. Yang, "Inferring taxi status using gps trajectories," arXiv preprint arXiv:1205.4378, 2012.

[59] L. Tang, X. Yang, Z. Kan, and Q. Li, "Lane-level road information mining from vehicle GPS trajectories based on Naïve Bayesian Classification," ISPRS International Journal of Geo-Information, vol. 4, pp. 2660-2680, 2015.

[60] L. Wang, X. Geng, J. Ke, C. Peng, X. Ma, D. Zhang, et al., "Ridesourcing Car Detection by Transfer Learning," arXiv preprint arXiv:1705.08409, 2017.

[61] C. C. Khor, S. Davila, W. B. Breunis, Y.-C. Lee, C. Shimizu, V. J. Wright, et al., "Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease," Nature genetics, vol. 43, p. 1241, 2011.

[62] M. Peng, J. Zhu, H. Wang, X. Li, Y. Zhang, X. Zhang, et al., "Mining event-oriented topics in microblog stream with unsupervised multi-view hierarchical embedding," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 12, p. 38, 2018.

[63] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, and Y. Zhang, "Arrhythmias classification by integrating stacked bidirectional LSTM and two-dimensional CNN," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2019, pp. 136-149.

[64] X. Niu, Y. Zhu, and X. Zhang, "DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces," in Global Communications Conference (GLOBECOM), 2014 IEEE, 2014, pp. 2745-2750.

[65] J. He, H. Chen, Y. Chen, X. Tang, and Y. Zou, "Diverse Visualization Techniques and Methods of Moving-Object-Trajectory Data: A Review," ISPRS International Journal of Geo-Information, vol. 8, p. 63, 2019.

[66] T. Feng and H. J. Timmermans, "Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data," Transportation Planning and Technology, vol. 39, pp. 180-194, 2016.

[67] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," IEEE Transactions on Visualization and Computer Graphics, vol. 19, pp. 2159-2168, 2013.

[68] X. Huang, Y. Zhao, C. Ma, J. Yang, X. Ye, and C. Zhang, "TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," IEEE Transactions on Visualization and Computer Graphics, vol. 22, pp. 160-169, 2016.

[69] W. Chen, F. Guo, and F.-Y. Wang, "A survey of traffic data visualization," IEEE Transactions on Intelligent Transportation Systems, vol. 16, pp. 2970-2984, 2015.

[70] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," IEEE Transactions on Visualization and Computer Graphics, vol. 19, pp. 2149-2158, 2013.

[71] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, "A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes," in Computer Graphics Forum, 2017, pp. 36-61.

[72] G. McArdle, A. Tahir, and M. Bertolotto, "Interpreting map usage patterns using geovisual analytics and spatio-temporal clustering," International Journal of Digital Earth, vol. 8, pp. 599-622, 2015.

[73] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, 2016.

[74] J. Klein and M. Moeschberger, "Survival analysis: Techniques for censored and truncated regression," ed: New York, NY: Springer-Verlag, 1997.

[75] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests," The annals of applied statistics, vol. 2, pp. 841-860, 2008.