ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 35 (2014) 720 – 729
Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Building knowledge from social networks on what is important to drivers in constrained road infrastructure

Panraphee Raphiphan [a,*], Arkady Zaslavsky [b,c], Maria Indrawan-Santiago [a]

[a] *Monash University, 900 Dandenong Rd, Caulfield East, VIC 3145, Australia*
[b] *CSIRO, PO Box 312, Clayton South, Vic, 3169, AUSTRALIA*
[c] *St.Petersburg National Research University of ITMO, Russia*

Abstract

The disproportionate growth of the number of vehicles compared to the available and even growing road infrastructure results in severe traffic congestion in metropolitan areas, causing tremendous tangible and consequential losses in all sectors, especially in fast developing Asian economies. Bangkok is one of the cities where traffic congestion is a crucial problem. Bangkok is an old city where high density residential areas are combined with ineffective road network. The Traffic Information Systems (TISs) can play a significant role towards improving traffic congestion problems. In this paper, we present the analytical results of our quantitative research which studies various aspects related to the knowledge-based TIS. The analysis of factors that affect Bangkok's traffic as perceived by drivers, or what we call influential factors (IF), along with the impact level of each IF are reported. In addition, the potential use of social networks for TIS is also discussed. This paper also highlights the success of using the social network to reach out the massive number of people who provide feedback and thus dramatically increase the usefulness of this information for the TIS. The reported results not only strongly confirm our selection of influential context attributes in our previous study, but also confirm the feasibility of using traffic-related data from social networks in Bangkok as context attributes for our framework. The Weight Mean Score method of influential factors presented in this paper can be further enhanced and become the metric to improve our proposed framework. Our analysis and result can also guide the design of the knowledge-based Traffic Information Systems. Although the research study was done based on Bangkok data, it can also be applicable to other cities that have similar road infrastructure problems as Bangkok.

Peer-review under responsibility of KES International.
Keywords: Intelligent Transportation System; ITS; Traffic Information System; context-aware; Bangkok; Intelligent System; Social Network

* Corresponding author.
E-mail address: panraphee.raphiphan@monash.edu

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.154

1. Introduction

Traffic congestion has always been a serious problem for commuters in the metropolitan areas of the country with constrained roads infrastructure such as Thailand, Indonesia, and Philippines. It causes unpredictable delays and increased cost of wasted fuel as well as inefficient use of other resources. Bangkok, Thailand is one of the cities where traffic congestion is a crucial problem caused by disproportionate growth of vehicles and roads and lack of discipline of road users. In addition, Bangkok is a very old city which has been growing in disorderly way. The town plan is hard to be assigned properly. The lack of well-planned traffic management and infrastructure also causes many problems. In 2012, Bangkok was voted to be the most traffic jammed city in the world. Bangkok's traffic problems have been getting worse since the government introduced a policy to refund tax for first-time car buyers. This policy has resulted in five million vehicles in a city which can only cope with less than two million cars[1]. The traffic problem causes tremendous tangible and intangible losses both economically and for quality of life.

In order to solve the traffic problem, Intelligent Transportation System (ITS) which refers to a system which integrates existing and emerging technologies with both information technology and telecommunication for facilitating the traffic system has been adopted in many countries, including Thailand, though with different purposes. Knowing the traffic condition in advance helps the drivers plan their trips more efficiently. Obviously, the Traffic Information Systems (TISs) can play a significant role towards improving the traffic problem. The goal of the TIS is to provide travelers with useful traffic information to assist their route choice decisions. Normally, the TIS involves network infrastructures and sensors to collect necessary traffic data. Then the data is processed to produce useful travel information and guidance, and disseminate recommendations to users. Polsawat, Siriwong Na Ayutaya, and Pattara-Atikom[2] have studied the impact of Intelligent Traffic Information Systems on Congestion Saving in Bangkok and reported that the traveler in Bangkok can benefit from Intelligent Traffic Information System with time saving during heavy traffic period.

Real time traffic congestion degree is one type of useful information to assist drivers in decision making. The congestion degree can be usually calculated on the basis of road side sensory data or mobile probes data. However, it is possible that the sensory data can be lost due to potentially unreliable communication, faulty sensors, or intermittently available mobile sensors leading to loss of important traffic data. The framework of adaptive context-aware traffic congestion estimation to compensate missing sensory data in Bangkok was proposed in our previous publications[3,4]. Our approach can overcome the uncertainty of roadside sensors or mobile sensor data. Our framework also enables the availability of traffic information at all times even though traffic sensory data might be lost through the adaptability of our algorithm based on acquirable context attribute instead of relying on only traffic sensory data. The experiments carried out using our proposed framework gave very promising results. However, we are pursuing further research on extending the framework by focusing on the influential factors which impact traffic congestion in Bangkok (which will subsequently become context attributes in our approach). Such research is presented in this paper.

The advent of online social networks has been one of the most exciting events in this decade. Many online social networks such as Twitter, LinkedIn, Facebook, and google+ have become increasingly popular. Such social networks are extremely rich in content, and they typically contain a tremendous amount of content which can be leveraged for analysis. Many approaches to utilize social network in traffic report have been proposed in recent years. The Thailand National Electronics and Computer Technology Center has created its own Twitter accounts named @traffy to convey traffic updates to and receive information from others[5]. Twitter has also been used by TMC Polda Metro Jaya Indonesia to spread the traffic news. The natural language processing was used to extract traffic information from tweets and reflect those data onto the map[6]. In addition, an attempt to harvest road traffic information from virtually indeterminate sources of data on Twitter timelines for real time mapping was done in[7].

We also proposed that our approach can be strengthened by aggregating more context such as traffic context exchanging through social network sites to compute the confidence factor[8]. Even though some groups have reported the approach to extract traffic information from context distributed through social networks, those works are not studied based on Bangkok road users and the system improvement according to the current change of environment and factors still remain to explore. We then pursue the research on the feasibility to utilize traffic data generated from social network's users for TIS in Bangkok as illustrated in section 5.

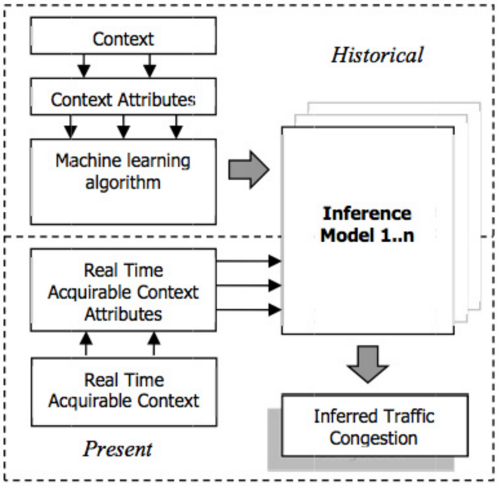
From the result analysis presented in this paper, we can confirm our initially selected influential context attributes along with the impact level of each context attribute. In addition, we proposed that our framework presented in[8] can be used as the less expensive alternative tool for making decisions on which influential context attribute is worth to concentrate. The research results presented in this paper does confirm our previous conclusions. We also study the potential use of social networks for TIS in Bangkok to improve the efficiency of traffic estimation system.

The remainder of this paper is structured as follows: Section 2 briefly describes our Context-Aware Traffic Congestion Estimation to Compensate Missing Sensory Data in Bangkok. Section 3 explains the methodology while the influential factor analysis and discussion can be found in section 4. The analysis and suggestion of potential use of social network to Traffic Information Systems (TIS) is shown in section 5. Finally, section 6 concludes this paper by summarizing our work, the contribution of our study, and the future work.

2. Context-aware traffic congestion estimation to compensate missing sensory data in Bangkok

In typical traffic dissemination systems, the traffic data is provided from the traffic sensors. However, the sensory data can be missing at particular time due to unsuccessful data transmission or bad weather condition, or the sensors can even be broken. In addition, if the system relies on mobile sensors which can be moved at any time, the sensory data can be intermittently available. Loosing sensory data can make the traffic report incomplete. We address this problem by providing the approach to approximate the missing sensory data by using machine learning techniques along with the context awareness paradigm. Our proposed framework aims not to rely on traffic sensory data as it is not available, but utilizes available *real time acquirable context* instead. Our approach is cost efficient and can fulfills the concept of “anytime”, “anywhere”, and “anything” for ubiquitous ITS and can achieve the availability, transparency, seamlessness, and awareness goals of ubiquitous services. Our proposed framework can handle uncertainty through the adaptation capability. The system can adapt to choose inference model according to the list of available context in real time. The detail of our context-aware traffic congestion estimation to compensate missing sensory data can be seen in our previous publications[3,4,8].

The process of our context-aware traffic congestion estimation to compensate missing sensory data framework starts by generating 1 to n inference models suitable for different sets of context attributes by machine learning algorithm. Each model is created for each set of real time acquirable context attributes. Each set compose of the possible combination of acquirable context attributes in real time. The model building process must be repeated n time until finish creating n models for n context attribute sets. If we have n influential context attributes, we will get $\sum_{k=1}^{n} {}^{n}C_{k}$ models for all possible combinations of context attributes where n is the total number of influential context attributes and k is the number of member(s) in each context attribute set. In real time inferring phase, if the system detect that the sensory data is missing at time t, a suitable inference model will be then chosen depending on available context at that time. Once a suitable model is selected, the real time acquirable context attributes are used as input for a selected inference model to generate an *inferred traffic congestion degree*. The overall process and algorithm are summarized in Fig. 1.




```
Procedure InferCongDegree
/*CtxAtrb is a set of current available context attributes */
Begin
/* Collect available current context */
AllContext := current available context

/* Extract context attribute by function "AtrbExtracter" */
CtxAtrb := AtrbExtracter (AllContext)

/* Filter only influenced context attributes by function
"FilterInfluencedCtx" */
CtxSet := FilterInfluencedCtx(CtxAtrb)
        If CtxSet match attribute in InferenceModel1
        Then Call InferenceModel1 (CtxSet,InferDegree)
                .....
                .....
        Else If CtxSet match attribute in InferenceModelN
        Then Call InferenceModelN (CtxSet,InferDegree)
        End if
/*output the inferred congestion degree*/
output InferredDegree
End
```

Fig. 1. (a) the overall process of our context-aware traffic congestion estimation to overcome uncertainty of sensory data ; (b) The algorithm for traffic congestion degree inference in real time.

The context attributes in our previously proposed framework[3,4,8] were initially chosen based on asking expert and based on opinion of researchers who work on this project. The context attributes we applied were *traffic congestion of adjacent road segment*, *rain level (T, S, M, H, VH, N/A)*, *day of the week (D1, D2,…, D7), time period of the day (p1, p2, p3,…p24), and school break (close, open)*. The traffic data used for our experiment are real traffic data log of roads in Bangkok administered by the National Electronics and Computer Technology Center of Thailand and the weather log used is from the Thai Meteorological Department.

The experiment on our proposed framework gave very promising result. However, we are pursuing further research on improving the framework by studying on the influential factors (IF) to the traffic congestion in Bangkok (which will become the context attributes in our approach). Our research on IF based on user's perception along with analysis is presented in this paper.

We also proposed that our approach can be improved by aggregating more context such as traffic context exchanging through social network to compute the confidence factor[8]. We thus pursue the research on potential use of social network for traffic information system in Bangkok as can be seen in section 5.

3. Methodology

The research methodology adopted in this study is quantitative research via web-based survey. The major reasons for using an online survey over the traditional survey are decreased costs, faster response-times, and higher quality for data-collection[9,10]. Regarding to[11], "Intercept Survey" was selected as sampling method for this study. The carefully designed questionnaire was pilot tested before using and found that it had a Cronbach's alpha of 0.84 for internal-consistency reliability.

The link of the questionnaire page was shared during December 2012 to February 2013 through websites, web boards, and the social network of the main author who has over 400,000 followers on her Twitter, Facebook and Google+. Any person who volunteers to help can click at the link and start answering the questions with no obligation. It is significant that the link of the online questionnaire page was shared through the social network such as twitter, Facebook, Google+ via the main author's account who has over 400,000 followers on her social media to stimulate the number of responses. This results in ability to receive 3,037 responses in short period of time. This emphasizes the success of using social network to reach out massive numbers of people.

The questionnaire consists of 20 questions in total which are categorized into 6 parts while the first part, screening question, is used to screen only participants who drive vehicles in Bangkok in routine manner in order to provide precise information. However, only selected results related to the scope of this paper are shown in this paper.

After the participants complete the online questionnaire, data file of 3,037 respondents was generated as the SPSS compatible format. Only the completed records with no error and pass the criteria of screening question were selected to be analyzed. In total, there are 861 respondents out of 3,037 records remain for analysis.

Referring to the number of driving licenses and transport personnel licenses as of 31 March 2014 in Bangkok reported by the Department of Land Transport of Thailand[12], the number of all driver licenses in Bangkok is 4,427,298. This number can be considered as the population size (*N*) for our research. According to the determining the sample size for research activities written in[13], our sample sizes (*S* = 861) which is the number of respondents who drive in routine manner in Bangkok (who pass the screening question) and complete the survey is enough. Answers from participants are then transformed into code and was analyzed by SPSS and some statistical methods.

4. Influential Factors (IF) to Bangkok traffic congestion

To improve our previously proposed context-aware traffic congestion estimation to compensate missing sensory data framework, we do a further research on discovering what could be the influential factors and how much each factor affects to the traffic in Bangkok. We asked the question "to what extent do you think the following factors affect the traffic in Bangkok?". The respondents answered by choosing the score from 1 to 5 while 1 represents "no-effect" and 5 represents "extreme effect".

4.1. Influential Factors (IF) analysis

In our IF analysis, we use the class interval technique. There are 7 factors which are Specific timing of the day, Specific day of the week, Group of the day, Public holiday, Density of vehicles on the adjacent roads, Rain Level, and Other incidents) such as accidents, road constructions, etc.). The Level of Importance are 1 = No effect, 2 = A bit effect, 3 = Medium effect, 4= A lot effect, and 5 = Extreme effect.

By applying the Weight Mean Score (WMS) method[14]

$$WMS = \frac{5F_5 + 4F_4 + 3F_3 + 2F_2 + 1F_1}{TNR}$$

When, WMS = Weight Mean Score, F5 = the number of "Extreme effect", F4 = the number of "A lot effect", F3 = the number of "Medium effect", F2 = the number of "A bit effect", F1 = the number of "No effect" and TNR = Total number of respondents

The class interval calculation is based on the following formula (when, *k* = number of classes).

$$\text{Class interval} = \frac{x_{max} - x_{min}}{k} = \frac{5-1}{5} = 0.80$$

Then, the criteria for interpreting the finding are 4.21-5.00 = Extreme effect; 3.41-4.20 = A lot; 2.61-3.4 = Medium; 1.81-2.60 = A bit; 1.00-1.80 = No effect.

Table 1. Descriptive statistic of IF analysis.

	N	WMS	Std. Deviation	Interpretation
Specific timing of the day (eg. 6-7 am, 11-12am, ..)	861	4.36	0.846	**Extreme effect**
Specific day of the week (eg. Mon, Tue, Wed, ..)	861	3.82	0.969	**A lot**
Group of the day (eg. Weekday / Weekend)	861	3.80	1.004	**A lot**
Public holiday	861	3.85	1.141	**A lot**
Density of vehicles on the adjacent roads	861	3.98	0.958	**A lot**
Rain Level	861	4.30	0.845	**Extreme effect**
Other incidents (eg. Road construction, accident, protest rally)	861	4.36	0.857	**Extreme effect**

From the analysis result shown in Table 1, we can group 7 factors into 2 groups which are

Group1: the factor that has a lot effect to the Bangkok road traffic which are *day, group of the day, public holiday,* and *the density of vehicles on the adjacent roads*.

Group2: the factor that has extreme effect to the Bangkok road traffic which are *time, rain level,* and *incidents.*

4.2. Correlation analysis

To study more on the correlation among each factor, we do the correlation analysis using Pearson Correlation. Only selected factors are chosen to be reported in this paper due to the limitation of space.

Table 2. Correlation analysis of some selected variables.

		Specific timing of the day	Specific day of the week	Density of vehicles on the adjacent roads	Rain Level
Specific timing of the day	Pearson Correlation	1	**.455****	.226**	.244**
	Sig. (2-tailed)		.000	.000	.000
	N	861	861	861	861
Specific day of the week	Pearson Correlation	**.455****	1	.276**	.185**
	Sig. (2-tailed)	.000		.000	.000
	N	861	861	861	861
Density of vehicles on the adjacent roads	Pearson Correlation	.226**	.276**	1	.336**
	Sig. (2-tailed)	.000	.000		.000
	N	861	861	861	861
** Correlation is significant at the 0.01 level (2-tailed)					

Using Pearson correlation, we test the null hypothesis, H_0, that there is no correlation in the population against the alternative hypothesis, H_1, that there is correlation;

$$H0 = \rho_{time,\, day} = 0 \qquad and \qquad H1 = \rho_{time,\, day} \neq 0$$

The range of correlation coefficient (r) conducted by[15] was applied. It can be seen in Table 2 that value of 0.455 is a positive correlation which means sound significant correlation between *time* and *day* variable. Therefore, reject *H0* at $\alpha = 0.01$. This can confirm that the *day* and *time* variable has high correlation. Thus combining *day* and *time* as context attributes in one context attribute set is suitable.

From the simulation result in[8] we suggested that the context attribute set that compose of only *day* and *time* or any context attribute sets that consist of these two variables can give high accuracy of obtained inferred traffic congestion degree. This analysis result is consistent with our conclusion in the previous proposed framework.

Furthermore, the analyzed result in this section can assist us to specify the factor that influent to the Bangkok's traffic, or what we call "influential factor" (IF). The result confirms our initially selection of influential context attribute in our previous work[3,4]. In addition, the result can let us specify how important of each factor to the traffic condition in Bangkok which also support what we concluded in[8]. This is useful for making a decision on which context attribute suitable to invest on especially under the resource constrained circumstance. The WMS of each IF can be further processed to become the weight for machine learning algorithm to improve our previously proposed framework in the next stage of our research.

Moreover, it is noticeable that the "incident" is in the *extreme effect* group. We are convinced that including the incident as context attributes should improve the accuracy of traffic congestion degree computation which can be the future work to improve the TIS.

5. The potential use of social network to traffic information systems

Even though there are some groups reported the utilization of messages distributed through social network and some groups have proposed methods to harvest traffic information from social network, those works are not studied based on Bangkok's users. We thus design the survey to discover how Bangkok road users use the social network and the potential of using it to provide the useful traffic data in Bangkok to improve existing system.

Table 3. the frequency use of social network * how often of mentioning about traffic condition cross tabulation.

			Mention_traffic_cond				Total
			Rarely	Not sure	Many time	Often	
Use Social Network	Rarely	Count	30	8	8	4	50
		% within Use Social Network	60.0%	16.0%	16.0%	8.0%	100.0%
		% within Mention_traffic_cond	6.5%	12.7%	3.3%	4.3%	5.8%
		% of Total	3.5%	.9%	.9%	.5%	5.8%
	1-3 days/week	Count	39	6	7	3	55
		% within Use Social Network	70.9%	10.9%	12.7%	5.5%	100.0%
		% within Mention_traffic_cond	8.4%	9.5%	2.9%	3.2%	6.4%
		% of Total	4.5%	.7%	.8%	.3%	6.4%
	4-6 days/week	Count	69	8	13	2	92
		% within Use Social Network	75.0%	8.7%	14.1%	2.2%	100.0%
		% within Mention_traffic_cond	14.9%	12.7%	5.4%	2.1%	10.7%
		% of Total	8.0%	.9%	1.5%	.2%	10.7%
	Everyday	Count	324	41	214	85	664
		% within Use Social Network	48.8%	6.2%	32.2%	12.8%	100.0%
		% within Mention_traffic_cond	70.1%	65.1%	88.4%	90.4%	77.1%
		% of Total	37.6%	4.8%	24.9%	9.9%	77.1%
Total		Count	462	63	242	94	861
		% within Use Social Network	53.7%	7.3%	28.1%	10.9%	100.0%
		% within Mention_traffic_cond	100.0	100.0%	100.0%	100.0%	100.0%
		% of Total	53.7%	7.3%	28.1%	10.9%	100.0%

We present our result analysis in the form of cross table in Table 3 and Table 4 to facilitate the brief interpretation and show the relations between variables. We were interested in the frequency use of social network of Bangkok road users, how often and what context they mention about traffic condition in social network. The analysis result turns out that it is feasible to take advantage of a significant number of social network users in Bangkok for traffic information system.

Table 3 describes the relation between frequency use of social network of Bangkok road users and how often they mention about traffic condition over social network. The analysis is discussed in later section.

Table 4 describes the relation between how often of mentioning traffic condition and what context included when mentioning about traffic condition of Bangkok road users.

Table 4. How often of mentioning traffic condition * what context included when mentioning about traffic condition (how do you mention?) cross tabulation.

			How do you mention?				
			Only the traffic cond.	Traffic condition and name of roads	Traffic condition and time	Traffic cond. and name of roads+time	Total
Mention about traffic condition	Rarely	Count	109	162	70	121	462
		% within Mention_traffic_cond	23.6%	35.1%	15.2%	26.2%	100.0%
		% within How do you mention?	63.7%	55.9%	53.0%	45.1%	53.7%
		% of Total	12.7%	18.8%	8.1%	14.1%	53.7%
	Not sure	Count	13	23	11	16	63
		% within Mention_traffic_cond	20.6%	36.5%	17.5%	25.4%	100.0%
		% within How do you mention?	7.6%	7.9%	8.3%	6.0%	7.3%
		% of Total	1.5%	2.7%	1.3%	1.9%	7.3%
	Many time	Count	37	78	38	89	242
		% within Mention_traffic_cond	15.3%	32.2%	15.7%	36.8%	100.0%
		% within How do you mention?	21.6%	26.9%	28.8%	33.2%	28.1%
		% of Total	4.3%	9.1%	4.4%	10.3%	28.1%
	Often	Count	12	27	13	42	94
		% within Mention_traffic_cond	12.8%	28.7%	13.8%	44.7%	100.0%
		% within How do you mention?	7.0%	9.3%	9.8%	15.7%	10.9%
		% of Total	1.4%	3.1%	1.5%	4.9%	10.9%
Total		Count	171	290	132	268	861
		% within Mention_traffic_cond	19.9%	33.7%	15.3%	31.1%	100.0%
		% within How do you mention?	100.0	100.0%	100.0%	100.0%	100.0%
		% of Total	19.9%	33.7%	15.3%	31.1%	100.0%

Discussion:

From our survey result, we can indicate that currently it is feasible to gain benefit from enough number of social network users. The majority of Bangkok-drivers spend time on social network regularly in everyday life (77.1%). Although about 54% of Bangkok road users hardly mention about traffic condition, nearly 40% of them mention quite often (often 10.9% and many times 28.1%). It is impressed that about 33% of those who mention about traffic condition include the location in their posts.

From Table 3, it is noticeable that 28.1% of road users who normally use social network mention about traffic condition many times. Furthermore, 48.4% use social network every day and rarely mention about traffic condition over social network.

From Table 4, it is interesting that if the respondents regularly mention about traffic condition over social network, most of them will include *the traffic condition, location,* and *time* (36.8%). It is useful to extract the location information when users include the name of the roads or the places such traffic condition occur in their messages. This result proves that we can obtain both traffic condition and location information from social network account of Bangkok road users. Consequently, the traffic data harvested from social network can be taken as one of the influential context attributes for our adaptive real time traffic congestion estimation system including other traffic estimation systems to enhance the competency and accuracy.

One possible approach to extract traffic data from social network is to create specific social network account such as twitter account to report and receive the traffic information from distributed users. This method give more simplicity of process than collecting all unrelated context distributed over all social networks because most of the content participated in specially created social network account shall be only traffic information. However, it is important that the more users engaged the more useful traffic data to be contributed to the system. Therefore, we must promote such account to have a great number of engaged users when deploying this approach. It is noteworthy that we could receive sizable feedback in short period of time after we distribute our online questionnaire through social network accounts which have over 400,000 followers. Therefore it is significant to let numerous users participated if we decide to use social network as part of traffic report services.

However, there are some limitations when using traffic context from social network. Firstly, the obtained traffic information from social network may be obsoleted and inaccurate because it might not be posted in the exact moment and place such traffic condition occurs. Secondly, the traffic information may only be reliable in short amount of time and becomes obsolete soon after it is posted. Thirdly, the creditability of information exchanging through the social networks is low. Fourthly, the traffic data posted on social network will be available intermittently for particular roads. Finally, most of the messages about traffic information posted over social media are ungrammatical[16]. The natural language processing, syntactic analysis, and social media data mining techniques should be applied[17].

According to some limitations stated above, from our point of view, we consider that the traffic information extracted from any social network sites is more suitable to be used as the supporting factor. It shall be used as the context attribute for our adaptive real time traffic congestion estimation system to compensate missing sensory data when other context attributes cannot be acquired in real time. Alternatively, it can be used to calculate the confidence factor to assist in making decision after traffic congestion degree has calculated rather than relying on it as the main input for traffic estimation.

6. Conclusion and future research

The Intelligent Traffic Information Systems can play an important role towards alleviating the traffic congestion. The framework of adaptive context-aware traffic congestion estimation to compensate missing sensory data in Bangkok that can overcome the uncertainty of traffic sensory data based on acquirable context attributes instead of relying on only traffic sensory data was proposed in our previous works[3,4]. In this paper, we present the statistical analysis of our recent quantitative research contributed to the TIS. The analysis on the influential factors and the analysis of social network data that we pursue to improve our framework are summarized along with the suggestions. This paper also demonstrates the distinct and apparent success of using social networks to reach out to massive numbers of people which gives dramatic advantage to the TIS. For those who decide to utilize the social network for implementing the traffic report, it is noticeable that the more number of people are engaged, the more benefits will be brought to the traffic information systems.

The result of research presented in this paper does confirm our previous selection of influential context attributes. Furthermore, it helps us define how important each context attribute is and how does it align with our simulation based result. The WMS of each IF presented in this paper can be further processed to become the weight for machine learning algorithm to improve our previously proposed framework in the next stage of our research. The analysis and result can also guide the design of the knowledge-based Traffic Information Systems. In addition, our study on potential use of social network for building knowledge-based traffic information systems demonstrates that it is feasible to extract Bangkok's traffic data and the location of such traffic situation from context acquired from social networks. This data can be utilized for our proposed adaptive context aware traffic congestion estimation system to improve efficiency and accuracy of estimation in our future research. Moreover, the result of the study on feasibility of utilizing the traffic data from social network in Bangkok can be beneficial to other researchers or government sectors desire to implement the traffic information system of Bangkok.

Bangkok, Thailand has traffic congestion as the crucial problem caused by disproportionate growth of vehicles and roads and lack of discipline of road users. In addition, Bangkok is a very old city which has been growing disorderly resulting in improper city plan. The lack of well-planned traffic management and insufficient resources also cause the traffic problems. Even though our study presented in this paper was done based on Bangkok's data, it can also be applicable to other cities that have similar characteristics to Bangkok as stated.

References

1. *10 monster traffic jams from around the world*. 2012 30 September 2012; Available from: http://www.bbc.co.uk/news/magazine-19716687.
2. Poolsawat, A., K.S. Na Ayutaya, and W. Pattara-Atikom. *Impact of intelligent traffic information system on congestion saving in Bangkok*. in *Intelligent Transport Systems Telecommunications,(ITST),2009 9th International Conference on*. 2009.
3. Raphiphan, P., et al. *Overcoming uncertainty of roadside sensors with smart adaptive traffic congestion analysis system*. in *Intelligent Vehicles Symposium, 2009 IEEE*. 2009.
4. Raphiphan, P., et al. *Context Aware Traffic Congestion Estimation to Compensate Intermittently Available Mobile Sensors*. in *Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on*. 2009.
5. Traffy Twitter ; Available from: www.twitter.com/traffy.
6. Endarnoto, S.K., et al. *Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application*. in *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*. 2011.
7. Frippiat, D. and N. Marquis, *Web surveys in the social sciences: An overview.* Population (english edition), 2010. **65**(2): p. 285-311.
8. Raphiphan, P., et al. *Real time traffic congestion degree computation for minor sensorless roads using cost efficient context reasoning*. in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. 2010.
9. Wright, K.B., *Researching Internet‐based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services.* Journal of Computer‐Mediated Communication, 2005. **10**(3): p. 00-00.
10. Dillman, D.A., *Mail and internet surveys: The tailored design method*. Vol. 2. 2000: Wiley New York.
11. Couper, M.P., *Review: Web surveys: A review of issues and approaches.* The Public Opinion Quarterly, 2000. **64**(4): p. 464-494.
12. Transports Statistics Sub-Division, P.D.o.D.o.L.T.o.T., *The Number of Driving Licences and Transport Personnel Licences Classified by Type As of 31 March 2014*, 2014, Department of Land Transport of Thailand: Bangkok.
13. Krejcie, R.V. and D.W. Morgan, *Determining sample size for research activities.* Educational and psychological measurement, 1970. **30**(3): p. 607-610.
14. Yamane, T., *Statistics: an introductory analysis*1973, New York: New York : Harper & Row.
15. Li, F., *Relationship between EFL Learners' Belief and Learning Strategy Use by English Majors in Vocational Colleges.* Journal of Language Teaching & Research, 2010. **1**(6).
16. Kosala, R., E. Adi, and Steven, *Harvesting Real Time Traffic Information from Twitter.* Procedia Engineering, 2012. **50**(0): p. 1-11.
17. Aggarwal, C.C., *Social network data analytics*2011: New York : Springer 502.