2-2017

Crowdsensing and analyzing micro-event tweets for public transportation insights

Thoong HOANG

Pei Hua (XU Peihua) CHER
Singapore Management University, phcher@smu.edu.sg

Philips Kokoh PRASETYO
Singapore Management University, eplim@smu.edu.sg

Ee-Peng LIM
Singapore Management University, eplim@smu.edu.sg

Crowdsensing and Analyzing Micro-Event Tweets for Public Transportation Insights

Thong Hoang, Pei Hua Cher, Philips Kokoh Prasetyo, Ee-Peng Lim
School of Information Systems, Singapore Management University, Singapore
Email: vdthoang.2016@phdis.smu.edu.sg, phcher@smu.edu.sg, pprasetyo@smu.edu.sg, eplim@smu.edu.sg

***Abstract*—Efficient and commuter friendly public transportation system is a critical part of a thriving and sustainable city. As cities experience fast growing resident population, their public transportation systems will have to cope with more demands for improvements. In this paper, we propose a crowdsensing and analysis framework to gather and analyze realtime commuter feedback from Twitter. We perform a series of text mining tasks identifying those feedback comments capturing bus related micro-events; extracting relevant entities; and, predicting event and sentiment labels. We conduct a series of experiments involving more than 14K labeled tweets. The experiments show that incorporating domain knowledge or domain specific labeled data into text analysis methods improves the accuracies of the above tasks. We further apply the tasks on nearly 200M public tweets from Singapore over a six month period to show that interesting insights about bus services and bus events can be derived in a scalable manner.**

***Keywords*-transportation; information extraction; classification; sentiment analysis; crowdsensing, micro-events analysing.**

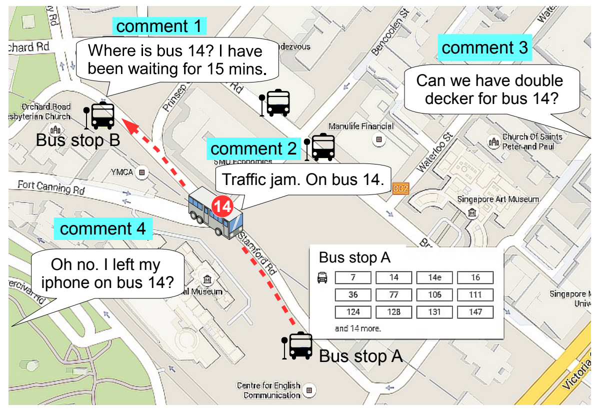


Figure 1: Example Scenario

I. Introduction

A. Motivation

Public transportation plays a critical role in both the economic and social activities in urban cities today. For example, in public bus transportation, commuters expect issues to be detected and rectified in a timely manner. In the past, feedbacks on public transportation were captured using hotlines and passenger surveys which requires significant efforts from both the passengers and transport operators. Although it is effective in sensing large events such as major accidents and breakdowns, it fails to capture granular events known as **micro-events**.

By micro-events, we mean spontaneous feedbacks from commuters on social media about public transportation services in buses/trains, at bus stops/train stations, and in other transport relevant contexts. In this paper, we focus on such micro-events from Twitter stream related to public buses. These events are small but important as they represent a large proportion of events experienced by commuters. Micro-events such as mentions of late arrival of buses and crowded buses reveal the quality of bus services. Major bus events are likely to be covered by many small micro-events too. Unfortunately, micro-events have largely been neglected in standard feedback systems. These micro-event feedbacks have not been collected and analyzed quick enough for timely response such as soliciting more detailed information about buses, or resolving some detected bus issues. For example, a micro-event reported a lost item left on a bus, the bus operator may immediately alert the bus driver to quickly retrieve the item.

Consider a scenario in Figure 1, bus 14 moves from bus stop A to bus stop B. At bus stop B, a commuter tweets about long waiting time. Another commuter mentioned traffic jam on bus 14 in Twitter. These two tweets can be associated with a bus stop and a moving bus respectively based on their geo-locations. There are also other non geo-coded tweets that mention bus 14. Geo-coded or not, these tweets or micro-events provide useful real-time feedback. The bus services at each bus stop, bus routes and other details provide the required contextual knowledge to analyze the tweets.

The micro-event feedback is different but complementary to: (a) GPS data from sensor systems installed in buses to track bus locations and arrivals at bus stops [1]; (b) GPS data from mobile phones of commuters involved in participatory sensing of bus locations [2]. Both (a) and (b) track bus locations and arrivals but could not capture commuters' comments. Hence, they could not tell the impact of bus events on commuters unless they are analyzed together with micro-event feedback, a topic we leave for future research.

B. Objectives

In this work, we propose a framework to crowdsense bus related micro-event feedbacks from Twitter data. We treat each bus related tweet as a micro-event which can be further analyzed for event type categorization, entity extraction and sentiment mining. We then develop and evaluate the methods for these analysis tasks and demonstrate the importance of domain knowledge and domain specific labeled data for these methods. We further apply our proposed crowdsensing and analysis methods on the 196 millions of Twitter posts generated by Singapore-based users to derive interesting insights about the public bus services in the Singapore city.

There are several research challenges in analyzing bus related micro-event tweets. First of all, micro-event tweets have to be identified from among numerous non-related ones in the Twitter stream. Secondly, the unstructured tweet content is not suited for quantitative sense-making. We need to introduce structures to the tweet content by extracting bus-relevant entities and opinions. Unfortunately, information extraction from tweet content is known to be a difficult task mainly due to laissez-faire and highly "contextualized" writing styles. The laissez-faire style gives rise to grammatically incorrect sentences, wrong spellings, and other problems usually not found in formal writing. The contextualized style assumes readers possess the appropriate domain, cultural and contextual knowledge to interpret content.

For example, the tweet: "`SBS7519B on Service 98 (SLBP 30)`" is clearly not proper sentence. It is unclear if NLP techniques are robust enough to handle such a sentence. The contextual knowledge about the bus plate number format (prefixed with `SBS` or `SMB`), dictionaries of bus service numbers and bus stops allows us to tell "`SBS7519B`", "`98`", and "`SLBP 30`" are bus plate number, bus service number and bus stop respectively. To interpret sentiments in bus tweets, localized slangs need to be identified. For example, the tweet "`Siao bus driver`" and "`Sui bus driver`" refer to totally different sentiments. They mean "crazy bus driver" (negative sentiment) and "nice bus driver" (positive sentiment) respectively. How to integrate the different tweet micro-events, interpret localized entities and sentiments, and analyze the patterns of public bus feedback are important research tasks to be addressed in this paper.

C. Contributions

We summarize our contributions as follows:

- We propose a sense-making framework for analyzing public transportation systems at the city scale using Twitter data. We use Singapore's public buses for illustration but the same framework is applicable to other public transportation systems in many other cities.
- Our approach gathers a very large Twitter dataset by crawling tweets generated by a large community of users based in Singapore, a city with 5.5 million population and extensive public bus system. From about 200M crawled tweets over a 6 month period, we determine the appropriate filter to select bus related tweets.
- We also develop context-specific entity extraction, event type classification, and sentiment mining methods on the bus related tweets using human labeled data. About 14,000 tweets have been manually labeled for training and evaluation purposes. Our experiments show that when using domain specific knowledge and labeled data, these methods deliver accurate results.
- We finally apply our proposed methods on a large Twitter dataset and empirically discover some interesting bus related micro-event findings. They include findings about event distribution, periodical patterns of events and sentiments, and correlation between event types. We also present some example cases to show the bus events discovered by combining our methods.

The main focus in this work is not to reinvent text mining methods but rather to integrate various existing state-of-the-art techniques together in a meaningful way to perform sense-making of for the public bus related tweets. Our sense-making results can be further combined with other data returned by sensor devices installed on buses and bus stops to provide even more contextual and offline knowledge to the commuter feedback.

II. Related Works

Collins et al. proposed to measure commuter satisfaction on Chicago's rail transit system by applying sentiment classification on 557 transit system related Twitter data [3]. While the work considers context-specific sentiment words for sentiment classification, it uses a simple aggregation of word level sentiments to derive tweet level sentiments instead of a machine learning approach. Furthermore, the work did not report the sentiment prediction accuracy.

Several works focus on detecting large events in public transportation systems. Sasaki et al. detected train disruptions using surges in tweet volume [4]. Limsopatham et al. studied 30 days of tweets in Glasgow to uncover temporal patterns in tweeting behavior of train disruptions [5]. They found that train disruption related tweets are often generated during rush hours on weekdays, and during late evenings on weekends.

In transportation specific micro-event analysis, Congosto et al. described the Metroaverias system that detects subway commuter complaints and train breakdown events in [6]. Complaints are classified into micro-event types (e.g., slowness, dirtiness, etc.) by specially crafted word dictionaries. The work however does not include the domain specific labeled data and machine learning techniques to improve the classification accuracy. Liu et al. also demonstrated the use of social media data to profile the gender of user population in the transportation systems [7].

This research, unlike the above studies, adopts an integrated approach towards analyzing the public transportation micro-events. We combine entity extraction ([8], [9]), event type classification and sentiment mining ([10], [11]) to obtain insights about the micro-events. These events may not cause surges in post volume and hence are not the target of bursty event detection research ([12], [13]). Throughout the research, we also evaluate the accuracy of different content mining components and ascertain their effectiveness and importance of using domain specific labeled data.

III. Framework Overview

As shown in Figure 2, our proposed crowdsensing and sense-making framework consists of three main group of components. The first group handles **data harvesting** from social media and the Web. Social media provides user generated content. The Web offers online data that can be crawled to construct a dictionary of transport related entities. Using a distributed Twitter data stream crawler, we are able to collect all public tweets generated by Singapore Twitter users. Tweet contents irrelevant to public transportation are removed. We also crawled and extracted transportation information from the web to build a transportation-specific dictionary.

The second group is the **sense-making engine** which transforms all unstructured tweet content into meaningful structured data. The key components are *entity extraction*, *event-type classification*, and *sentiment mining*. Entity extraction determines the transportation related objects involved in different micro-events which can be pre-defined or derived from data. Sentiment mining assigns a sentiment value (positive, neutral or negative) to every micro-event tweet.

Finally, **empirical analysis** can be conducted on the structured data generated by the sense-making components. The analysis includes *distribution analysis* where micro-event volume, entity, event, and sentiment distributions are analyzed to understand overall commuter feedback. *Periodical pattern analysis* focuses on observing periodical trends in the commuter feedback. Finally, *correlation analysis* seeks to determine how entity, event and sentiment may be correlated with one another in micro-event feedback.

IV. Data Collection

Crawling and Filtering: We utilize Twitter API to collect all tweets from about 150K Singapore Twitter users from 1 January 2015 to 30 June 2015. These users represent the more active Singapore users who post their content publicly. They are obtained by first selecting a few seed well-known Singapore users, followed by few rounds of snowball sampling of users through the follow links. From these sampled 150K users, we collected about 197M tweets during the data collection time period. We perform two additional filter steps to remove tweets which were not related to bus transportation. The first step filters away retweets as they do not contain new content compared with the original tweets. The second step selects tweets which contain the term "bus". This results in 139,108 relevant tweets which we call the *bus tweets*. This number amounts to about 700+

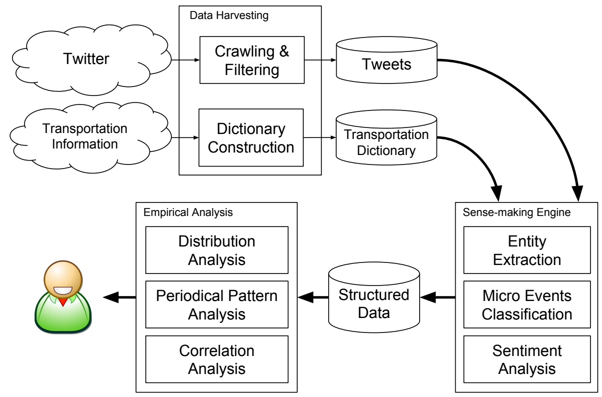


Figure 2: Sense-making Framework

tweets per day. While it is possible the filtering criteria misses out bus related tweets that do not contain the "bus" keyword, such a filter ensures that the precision of bus related tweets is very high. We randomly picked 1,000 tweets of the selected tweets and manually labeled them. It was found only 63 tweets that were not bus relevant suggesting that the *precision* (=∼94%) is indeed very high.

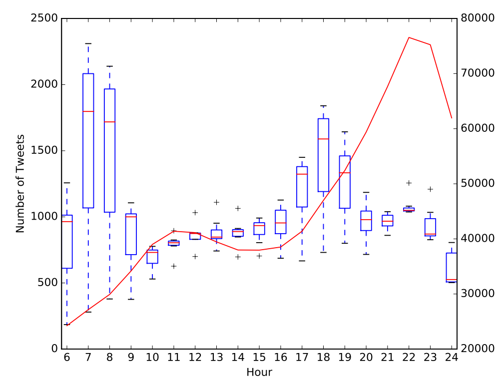


Figure 3: Boxplot of hourly distribution of bus-related tweets and hourly general tweets trend

The hourly distribution of bus tweets is shown as boxplot in Figure 3. The line in the figure shows the hourly trend of all tweets which always peaks right before midnight. The boxplot however shows that bus tweets are mostly generated during the morning and evening hours corresponding to the start and end of working days. It also suggests that the bus tweets effectively capture the micro-events when commuters use the public bus system.

Entity Dictionary Construction: We focus on three entity types: *bus service*, *road* and *bus stop*. In particular, we crawled all bus service numbers from *My Transportation* (`http://mytransport.sg/`), a website maintained by Singapore's government. We also crawled all road

Table I: Examples of regular expressions for extracting bus service number.

Regular expression	String matched
"*bus* $[0-9]+$"	Waiting for bus *911* at Woodlands interchange can kill me. :(
"*service* $[0-9]+$"	In service *135*.
"*no* $[0-9]+$"	no *966* is like a school bus every morning.

names in Singapore from `http://www.nearby.sg`, a location search service website. Finally, we crawled bus stop information including the bus stop names from *Electronic Guide* (`http://www.transitlink.com.sg/`). The constructed dictionary contains 352 bus service number, 3,910 road names and 4,743 bus stops which will serve as the domain knowledge for our subsequent analysis.

V. Entity Extraction

Methods. Entity extraction (EE) refers to the detection of bus service, road and bus stop in the text. A good survey of EE can be found in [14]. Our focus here is to combine the existing methodologies with domain specific knowledge so as to be able to derive more accurate extraction results. We study three different EE approaches, namely regular expression, classification and conditional random fields.

- **Regular Expression (RE)**: This approach uses manually crafted patterns to identify entities in a piece of text. When an entity in text can be matched with regular expressions created for different entity types, this entity is assigned the multiple matching types. Table I shows some regular expressions developed in this research. For bus service number, we check whether the word before number related to bus like "service", "bus", "no", etc. For road and bus stop, we also exploit the dictionary of road names and bus stop names to derive regular expressions. This results in 14, 4089 and 3909 regular expressions for extracting bus service numbers, roads and bus stop names respectively. Note that our RE are non case-sensitive.
- **Classification (CLF)**: The RE is domain specific but not always reliable, and therefore may not detect all entities in the tweets. For example, the bus service 858 can be easily extracted from the tweet "*@cyongqing go bus stop behind 888. Take bus 858.*" using RE since it has the word "bus" before the number 858. However, RE is unable to extract bus 888 as it is expressed in a phrase structure not captured by any defined regular expression patterns. RE also fails when entities are spelt wrongly. For example, the tweet "*bus 147 did not come to opp burlingtn sq*" contains bus stop name "*opp burlington sq*" misspelt as "*opp burlingtn sq*". To extract entities not covered by regular expressions, we can train classifiers using labeled entities in tweets, In this paper, we evaluate Naive Bayes, SVM and logistic regression classifiers.
- **Conditional Random Fields (CRF):** CRF [15] predicts the sequence of labels for an input sequence of words. Unlike the CLF approach which assumes entity labels in a tweet are independent from one another, CRF considers the ordering of these words and hence can potentially achieve better performance than CLF.

Features. We define a set of features for each word token in a tweet, and these features are used in classification methods and CRF to assign entity labels to the word token. The features (158 of them) are grouped into the following seven categories:

- *Dictionary:* Does the token match with bus service, road name or bus stop name?
- *Type:* Does the token start with a capital letter? Is the token a numeric, a word or a stop words?
- *Property:* Is the token a name-mention (i.e. starting with @), or a hashtag (i.e. starting with #)?
- *Part-of-speech:* Is the token a name, verb, noun, or assigned with other part of speech tags. To cope with the loose sentence structure in tweets, we use *tweetNLP* Part of Speech tagging package [16] and the set of tags returned by tweetNLP.
- *Regular expression:* Is the token determined to be a bus service number by any regular expression associated with bus service? Is the token determined to be a road name by any road name regular expression? Is the token determined to be a bus stop name by any bus stop name regular expression?
- *Before-word:* We define the features based on word token that comes before the target token. We first collect all tokens appearing before bus service, road name and bus stop name in the training data and call them the before-tokens of the respective entity types. The before-word features are then defined by: (i) whether the word token before the target token is among the before-tokens of each entity type, and (ii) whether the word token before target token is assigned each of the *dictionary*, *type*, *property*, *part-of-speech*, or *regular expression* features.
- *After-word:* These features are similar to before-token features except that the latter are derived from the words that comes after the target token.

Table II: F1 results of entity extraction.

Methods	Bus service	Road	Bus stop
Regular expression	0.81	0.75	0.61
Naive Bayes	0.89	0.74	0.63
SVM	0.93	0.77	0.64
Logistic regression	0.93	0.77	0.65
CRF	**0.95**	**0.79**	**0.68**

Experimental Results. Given that we have three entity extraction methods for three types of entities, we evaluate

Table III: Distribution of tweets with extracted entities

	Bus service	Road	Bus stop
Tweets	6054 (60%)	739 (6.8%)	4004 (39.6%)
Entities	277	186	829

the accuracy of all method-entity type combinations. We first need to construct a labeled tweet set for training CLF and CRF methods. We selected bus tweets containing some bus service number, and among these tweets further selected those with road name or bus stop name. Only 142 tweets meeting the selection criteria were presented to human anonotators for entity labeling. The annotators are Singapore users with some knowledge about the local bus system. The annotators assigned every word in the tweet with one of the four labels (i.e., none, bus service number, road name, and bus stop name). Only 6% of word tokens are labeled as road, 6.6 % of them are labeled as road names, and 6.5 % of these tokens are labeled as bus stops. Hence more than 80% of the word tokens in these tweets are not assigned any entity label. The F1 results of entity extraction using five-fold cross validation are shown in Table II. The results show that CRF outperforms the other two methods for all entity types. It also shows that the regular expression method already performs quite well. Another interesting finding is that the extraction of bus service numbers is easier than that of road and bus stop names. This may be caused by some road names being used as bus stop names and deciding between the two requires much contextual knowledge and human intelligence.

With the above results, we choose CRF to be the entity extraction method for *all* 139,108 bus tweets. Table III shows the number of tweets containing extracted entities of different types and the number of extracted entities. There are altogether 10,902 bus tweets containing extracted entities. This represents 7.8% of bus tweets. In other words, most users do not mention relevant entities in their bus tweets, e.g., "bus late 10mins" and "i am waiting at the bus stop". Most of the entity-mentioned bus tweets or micro-events mention bus service numbers (60%), followed by bus stop names (39.6%). Only 739 (6.8%) out of 10,902 tweets mention road names. The numbers of entity mentioning bus service number (227), road name (751) and bus stop name (4,051) are similar to the numbers of tweets. In other words, most tweets only mention one road and bus stop name.

VI. Micro-Events Classification

Event type labels. As we treat bus related tweets as micro-events, we want to categorize them into a number of event types to determine the popular bus related events shared by Singapore Twitter users. The automated classification of event types in turn enables many interesting applications, including routing micro-events to relevant authorities for follow-up actions, and for combining micro-events with on-board bus sensor data for bus monitoring. Unfortunately, no one knows the types of events until the tweets are manually judged. We therefore engaged a number of annotators to determine the appropriate event type labels for the 10,902 tweets containing some bus related entities found in our CRF entity extraction results. Ten relevant event types emerged from the manual labeling process as shown in Table IV.

Table IV: Event Types and Labeled Tweets

Event Type	Description	Count
Wait	Long wait at bus stop	859
Missing	Mention of someone having a lost item	158
Skip	Bus failing to stop at some bus stop	128
Slow	Slow moving traffic	117
Accident	Mention of some accident	109
Crowd	Mention of crowded environment	71
Bunching	Mention of bus bunching	66
Queue	Mention of long queues of people	49
Jam	Mention of traffic jam	11
Breakdown	Mention of some vehicle breakdown	7

Table IV shows the distribution of 10,902 tweets after manual labeling. Each tweet can be assigned zero, one or more event types. For example, the tweet "`Siol ah been waiting 30 mins for bus 106 all came all full cb or what`" has been assigned two event types, *wait* and *crowd*. Also, most tweets mention the "wait" event, which often occur when the commuters are at the bus stops complaining about waiting for some bus services. In our next analysis, we focus on the six most popular event types (i.e. *wait*, *missing*, *skip*, *slow*, *accident*, and *crowd*). Note that 9,356 of the tweets are not assigned any of the six event labels and we label them as *null-event*.

Classification methods. We address the event type classification problem as a binary classification task. We adopt a *one against one* (OAO) classification strategy to detect event types in Twitter. In particular, with six event types and one null-event type, we trained $C_7^2 = 21$ binary classifiers. Each classifier is trained using the labeled tweets of the corresponding pair of event types. To predict events on an unlabeled tweet, the event type predicted by most of the 21 classifiers is assigned to this tweet. If there is a tie, we assign the tweet the event type with the higher classification score. We also applied the *one against all* (OAA) classification strategy in which a classifier is trained for each event type against all other event types. We found the accuracy result of OAO is better than that of OAA possibly due to the very imbalanced labeled data distribution. We only have few labeled tweets per each event type, making the negative class much larger than positive class [17].

The binary classifiers can be based on different classification algorithms. We have tried out Naive Bayes, SVM and logistic regression. We also introduce an unsupervised *matching* baseline which detects some event word such as "missing" in a tweet for classifying the tweet under the

Table V: F1-scores of 10,902 tweets for micro events detection using NaiveBayes, Support Vector Machine (SVM), Logistic Regression(LR) and matching.

	NaiveBayes	SVM	LR	Matching
Accident	0.71	**0.84**	0.75	0.42
Crowd	0.05	**0.47**	0.31	0.28
Missing	0.58	**0.86**	0.81	0.04
Skip	0.54	0.73	**0.74**	0.59
Slow	0.11	**0.52**	0.41	0.33
Wait	0.54	**0.67**	**0.67**	0.18

Table VI: Number of tweets predicted for each event type.

Acc.	**Crowd**	**Miss.**	**Skip**	**Slow**	**Wait**
217	2944	493	419	4062	13716

missing event type, and vice versa for other event types. The term frequency of words in the tweet is used as a feature.

Experimental results. We applied five-fold stratified cross validation and obtained the F1 results as shown in Table V. In this task, SVM outperforms all other classification methods for all event types except *skip* and *wait*. LR outperforms SVM by small margin for these two event types.

We then apply the best method, i.e., SVM, to all bus tweets and obtain the predicted event type distribution as shown in Table VI. Similar to our labeled data, "wait" event has a highest number of tweets mentioned.

VII. Sentiment Analysis

Sentiment prediction method. In sentiment analysis, our goal is to determine the sentiment labels (i.e., *positive*, *neutral* and *negative*) of bus tweets. Given a tweet has only 140 characters, it is reasonable to explore natural language processing (NLP) sentiment mining approach. In this study, we use the classification method based on **Recursive Neural Tensor Network (RNTN)** [18], a state-of-the-art method with high prediction accuracy for movie review sentiment mining. This method predicts five sentiment labels (0-very negative, 1-negative, 2-neutral, 3-positive and 4-very positive). For our purpose, we combine the first two into negative sentiment and last two into positive sentiment.

In RNTN method, each sentence has to be represented in a binary tree (or sentence tree) in which every node is assigned a sentiment label[1] [19]. As we form the binary tree for a given tweet with multiple sentences, we concatenate the sentence specific binary trees together using a left-deep tree structure. We also modify the sentence tree construction to include emoticons and emojis which are common in tweets.

Labeled data. To train and evaluate our sentiment prediction methods, we manually label 3898 tweets randomly selected from bus tweets. The annotators live in Singapore

[1]This label is manually assigned during training and predicted when applying the trained RNTN.

Table VII: Ground truth sentiment labeled tweets.

Very neg.	Neg.	Neutral	Pos.	Very Pos.
712 (18.3%)	717 (18.4%)	1599 (41.0%)	587 (15.1%)	283 (7.3%)
1429		1599	870	

and are familiar with the local slang. With each tweet represented as a binary tree, an annotator assigns the sentiment labels of tree nodes in a bottom up manner, which is a laborious effort. For example , a tweet "`surprisingly bus 5 is rather empty`" may be manually labeled as a binary tree *(4 (3 surprisingly) (4 (2 (2 bus) (2 5)) (4 (2 is) (4 (2 rather) (4 empty)))))*. The distribution of tweet level sentiment labels is shown in Table VII.

Experimental Results. We evaluate RNTN method trained using labeled bus tweets, *RNTN(bus)*, against two baseline methods, namely RNTN method trained using movie reviews *RNTN(movie)* (which is the default method in the Stanford's package) and SVM trained using term frequencies. Table VIII shows the F1 results using stratified five-fold cross validation performed on the 3898 labeled tweets. RNTN(movie) performs very poorly due to the use of training data not from the bus application domain. SVM, based on bag of words, performs the best with F1=0.7 for both neutral and negative labels, and 0.66 for positive label. RNTN(bus) delivers the a comparative accuracy with F1 of 0.59, 0.66 and 0.71 for positive, neutral and negative labels, respectively. SVM performed better than RNTN(bus) due to small data size. The results also indicate that positive label is harder to predict, perhaps due to smaller number of tweets.

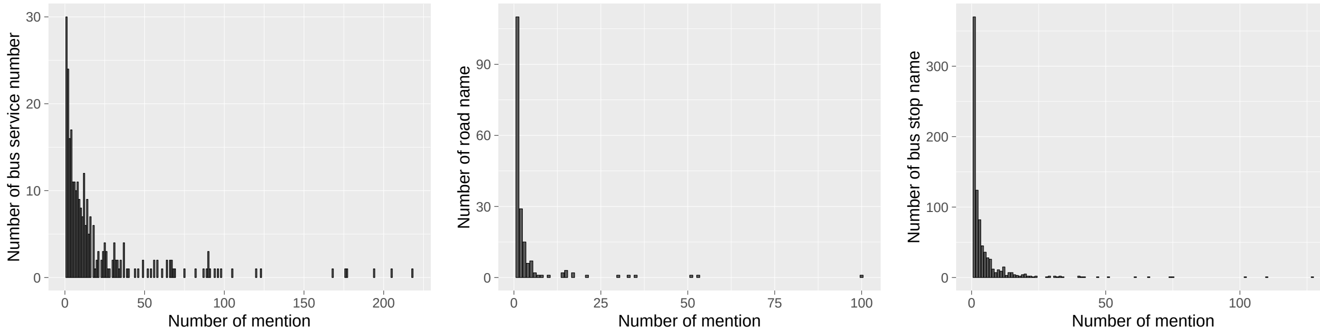
Comparison with emoji-based method. We are curious how the RNTN(bus)'s F1 results compared with emoji-based method which assigns the predicted sentiment by averaging the pre-defined sentiment score of emojis found in a tweet. The sentiment score for each emoji was extracted from the previously manually labeled ground truth tweets. Note that emoji-based method is only applicable to tweets with emoji(s). We evaluate using 700 manually labeled tweets with emojis that have not been included in the training data of RNTN(bus) method. The F1 results in Table IX shows that RNTN(bus) method still outperforms the emoji-based method. Hence, in our subsequent analysis, we use RNTN(bus) sentiment classification method in our empirical analysis. Emoji-based method did not perform well because emojis used, e.g. 😂, do not neccessarily reflect a clear nor consistent sentiment.

VIII. Empirical Analysis

We divide the analysis into distribution analysis, periodical trend analysis and correlation analysis.

A. Distribution Analysis

1) Entity mention distribution: First, we analyze the mention distribution of different entity types found in the



(a) Bus service number (b) Road name (c) Bus stop name

Figure 4: Mention distribution of entities.

Table VIII: F1 results of sentiment prediction.

	RNTN(movie)	**SVM**	RNTN(bus)
Positive	19%	**66%**	59%
Neutral	42%	**71%**	66%
Negative	61%	**74%**	71%

Table IX: F1 results for 700 test tweets.

	Emoji-based method	SVM	**RNTN(bus)**
Positive	38%	26%	**48%**
Neutral	49%	20%	**51%**
Negative	64%	56%	**77%**

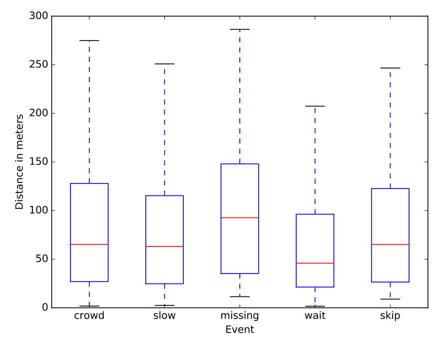


Figure 5: Distance of Event-mention Tweets to Nearest Bus Stops

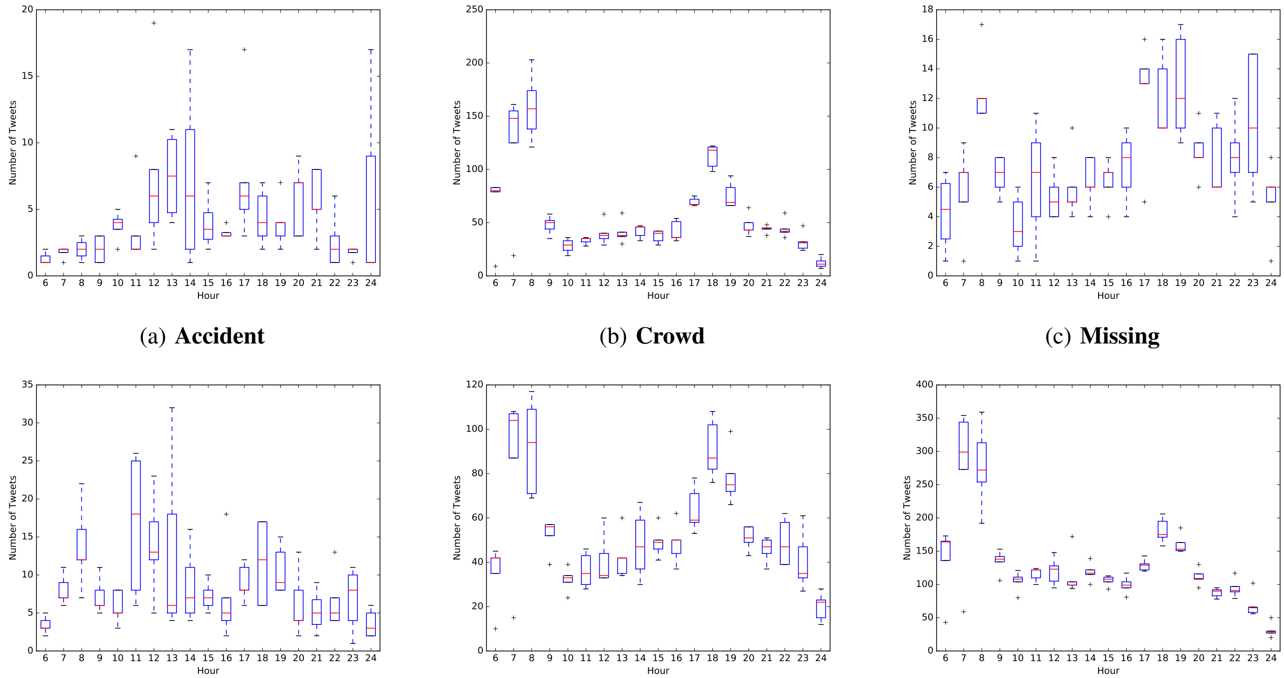
bus tweets. Figure 4 shows that among the bus tweets, bus service numbers are mentioned more often than road names and bus stop names. On the other hand, there are also many entities in our dictionary not mentioned at all. For example, 75 (21%) bus services, 3724 (95%) road names, and 3914 (83%) bus stop names in our dictionary are not mentioned in our 6-month bus tweets. The highly skewed distributions also suggest that only few entities account for most of the entity mentions. For example, the top 20% bus service numbers account for more than 80% of bus service mentions, following the Pareto Principle also known as the 80-20 rule. The same can be said for road name and bus stop name mentions.

2) Event tweet's distance from bus stop: We next examine a subset of geo-coded bus tweets that mention events and study how far these tweets are from the nearest bus stops. The intent is to see if tweets mentioning different event types are close to bus stops. There are 12,388 such tweets representing 8.9% of all bus tweets, a proportion much higher than the 2.8% geo-coded tweets for general tweets (the 200M tweets). We leave out "*accident*" event as the number of geo-coded bus tweets mentioning accidents is very small. As shown in Figure 5, "*wait*" tweets are closest to bus stops with median value below 50 meters. This is reasonable as they are likely generated by commuters at the bus stops. Tweets with in-bus events such as "*crowd*", "*slow*", and "*skip*" are farther away from bus stops because they are likely to be generated between bus stops. Tweets with "*missing*" events appear to be the furthest away from bus stops possibly because owners, who have lost their items, bring up the matter only after they have reached their final destinations.

B. *Periodical Pattern Analysis*

In this analysis, we focus on periodical patterns observed in the bus tweets. Such patterns help us to identify the daily event and sentiment rhythms of the city. Appropriate resources to address the bus micro-events can then be scheduled according to these rhythms.

1) Event periodical patterns: Figure 6 shows the hourly trend of event tweets during the weekdays. The figure shows that some event types (i.e. "*crowd*", "*slow*", or "*wait*") has high number of tweets during the morning peak hours (from 6 to 8am) and to a smaller extent during the evening peak hours. This is reasonable as most people use bus services during the peak hours and hence more likely to complain a lot about the bus services. The tweets mentioning event "*missing*" are usually occurred during evening time. We believe that most Twitter users only recognize a lost item after coming



(a) **Accident** (b) **Crowd** (c) **Missing**

(d) **Skip** (e) **Slow** (f) **Wait**

Figure 6: Hourly Trend on Weekday for each micro-event

back home, therefore they may not capture "missing" event in real time. The other events "*accident*" and "*skip*" do not observe the same periodical patterns. We also believe that these two events may occur when users share their news after reading them. For example, the tweet "*Accident involving a car and an SMRT bus. http://t.co/r2x9ooyS9U*" has the event "*accident*" as a user want share the accident news to her friends. Note that we leave out the results of weekend since they has similar trends as weekday.

2) Sentiment periodical patterns: We observe high number of positive and negative sentiment tweets between 7am and 8am as well as between 5pm and 7pm during weekdays as shown in Figures 7a and 7b. This again can be attributed to the rush hours commuters travel to work and return home. To determine if there is a change in the overall sentiment during the weekdays, we examine the proportion of tweets in Figure 7c. The figure shows that there is a slow increase of proportion of positive sentiments from morning to evening time. This suggests that people begin with a relatively less positive sentiment in the morning and become more positive in the later part of the day. This result is interestingly different from the Golder and Macy's results which say that people have better mood in the morning than in the evening [20]. Further works can be conducted to explain the possible differences. We also left out the weekend results since they have similar trends as weekday.

Table X: Number of bus services for event type pair.

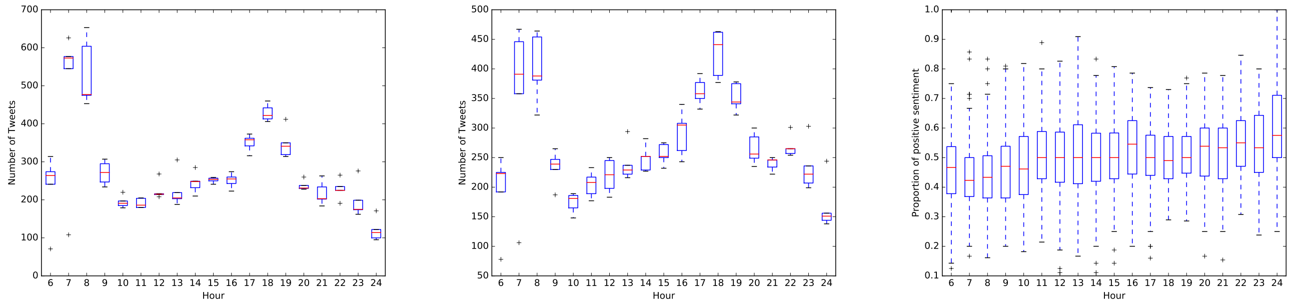
	crowd	*skip*	*slow*	*wait*
crowd	91	18	61	60
skip	-	33	20	25
slow	-	-	105	99
wait	-	-	-	184

C. Correlation Analysis

In this analysis, we examine the correlation between different entity and event types to determine if they co-occur in bus related tweets. We also want to know if there are some event types that co-occur among bus services.

1) Correlation between event types and entity types: Figure 8 depicts the number of entity mentions in bus related tweets about different event types. We observe that most of "*accident*" event tweets mention road names. The tweets of other event types, namely, "*crowd*", "*missing*", "*skip*", "*slow*", and "*wait*" (especially "*wait*"), mention bus service numbers mostly. This finding is reasonable as locations of accidents are usually of interest to users. The other event types are usually complaints directed at buses.

2) Events Correlation: We next study the number of bus services involving different pairs of event type. We focus on a subset of bus tweets that mention bus service numbers which we call the *bus service tweets*. The bus service tweets are further grouped by bus service numbers. Two event types



(a) **Negative Sentiment** (b) **Positive Sentiment** (c) **Proportion of Positive Sentiment**

Figure 7: (a) and (b) shows the weekday number of hourly negative and positive sentiment trends; and, (c) weekday proportion of positive hourly weekday trends.

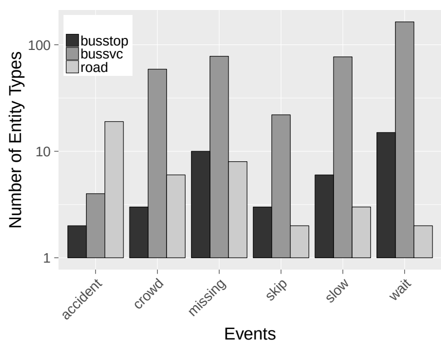


Figure 8: The correlation between micro-events and various entity types based on bus service number.

Table XI: χ^2 test of event type pairs ($\rho = 0.05$) .

	crowd	*skip*	*slow*	*wait*
crowd	-	$\rho < 0.05$	$\rho < 0.05$	$\rho < 0.05$
skip	-	-	$\rho > 0.05$	$\rho > 0.05$
slow	-	-	-	$\rho < 0.05$

co-occur in a bus service when both events are mentioned among the corresponding bus service tweets. Table X shows the number of bus services involving each event type pair. We find that "*crowd*" event type co-occurs many times with "slow" and "wait" events among bus services, and "slow" event co-occurs many times with "wait" event. Table XI presents the χ^2 test results which show significant correlations. Note that "accident" event has too few bus service numbers, and "missing" event is not related to bus service provider. Hence we exclude these events from this analysis.

D. Case Examples

In this section, we select two case examples with each example involves several micro-events detected for a single bus service on a specific date. The purpose is to demonstrate the efficacy of our sensemaking framework.

Table XII: Example Events

Time	User	Content (Predicted Event)
Example 1: Bus 15 on May 6, 2015		
8:08am	hea*	8.08 nope still no bus or humans Unless you count bus 15 as a bus (**wait**)
8:19am	hea*	8.17 tp shuttle bus can just go and eat sh*t la. I wait for like so long and still no sign. Bye bus 15 here I come (**wait**)
8:21am	hea*	Now bus 15 become express way alr. Wts (**wait**)
8:38am	hea*	8.37 nop all bus 15 is packed all become express bus. I am going to be late conform not chance (**crowd**)
9:19am	Meg*	do not trust bus 15 on late days....zzz hello infinite ESI (**wait**)
Example 2: Bus 182 on May 12, 2015		
10:51am	Stcom	SBS bus service 182, 182M to skip bus stops ... (**skip**)
11:05am	SGnews	[ST] SBS bus service 182, 182M to skip bus ... (**skip**)
11:05am	sgbroadcast	[ST] SBS bus service 182, 182M to skip bus ... (**skip**)
11:15am	spore88deal	SBS bus service 182, 182M to skip bus stops ... (**skip**)
11:15am	Sporecitybling	SBS bus service 182, 182M to skip bus stops ... (**skip**)
...	...	... (7 more tweets with similar content) ....

Note: User ids and partial content have been masked to protect user privacy and to censor vulgar language.

Example 1: As shown in Table XII, the first example happened on May 6, 2015 around 8am when users `hea*` and `Meg*` posted several tweets about waiting for Bus 15 and crowded Bus 15. Some of them have been correctly predicted with the wait event label.

Example 2: This is an event of bus services 182 and 182M rerouted to skip some bus stops shared by mostly Twitter accounts of news agencies and online news aggregators. They have been correctly predicted with the skip event label.

Example 1 is a unplanned event while Example 2 is a planned event. The capability to capture both unplanned and planned events is a new "discovery" of our sensemaking framework. It also illustrates the possibility to actually deploy the framework in real environment.

IX. Discussion & Conclusion

As social media becomes the de-facto platform for public content sharing, using it for crowdsensing micro-events relevant to public transportation and making sense out of these unstructured social media content will be the next trend. We have shown through sensing Twitter data at the city scale and using a combination of sensemaking methods, we can effectively harvest bus related tweets reporting micro-events in the bus system. We have also developed entity extraction, micro-event classification and sentiment analysis components adaptable to this application domain.

Our experiments show that the accuracy of these components benefit very significantly from using domain knowledge (e.g., regular expressions for extracting entities) and domain relevant labeled data (for sentiment analysis). The key findings of our empirical analysis on the gathered 140K bus tweets include: (a) bus tweets are mostly generated during morning and evening peak hours of weekdays; (b) few entities are responsible for most entity mentions in the bus tweets; (c) "*missing*" event tweets are found to be furthest away from bus stops while "*wait*" event tweets are closest to bus stops; and (d) most "crowd", "slow" and "wait" mentions of events occur during the morning and evening peak hours but not for "accident", "missing" and "skip" events. (e) the sentiment expressed in bus tweets are generally negative, and the proportion of positive sentiment increases gradually from morning to the evening during weekdays.

To our best knowledge, this work is the first that demonstrate an integrated approach to derive such findings at the city scale. This approach can be further applied to other cities as well as to other forms of public transportation, e.g., subways and taxis. This approach will thus facilitate the study of public transportation in different cities and countries making it possible to compare them. Beyond this, there are several interesting research directions including extracting actionable opinions from users for improving transportation services, and integrating the content analysis with bus/bus stop sensor data analysis. These methods have been incorporated into bus.sense, a realtime bus analytics system which aids the monitoring and response tracking of the bus micro-events (see bussense.org).

Acknowledgments. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

References

[1] K. J. Schmier and P. Freda, "Public transit vehicle arrival information system," 1999, uS Patent 6,006,159.

[2] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: predicting bus arrival time with mobile phone based participatory sensing," in *MobiSys*, 2012.

[3] C. Collins, S. Hasan, and S. V. Ukkusuri, "A novel transit rider satisfaction metric: rider sentiments measured from online social media data," *Journal of Public Transportation*, vol. 16, no. 2, 2013.

[4] S. Kenta, N. Shinichi, U. Koji, and C. Kenta, "Feasibility study on detection of transportation information exploiting twitter as a sensor," in *ICWSM*, 2012.

[5] L. Nut, M. Albakour, M. Craig, and O. Iadh, "Tweeting behaviour during train disruptions within a city," in *ICWSM*, 2015.

[6] M. Congosto, D. Lorenzo, and L. Sánchez, "Microbloggers as sensors for public transport breakdowns," *IEEE Internet Computing*, vol. 19, no. 6, pp. 18–25, 2015.

[7] W. Liu, F. Al Zamal, and D. Ruths, "Using social media to infer gender composition of commuter populations," in *ICWSM*, 2012.

[8] M. W. Berry and M. Castellanos, "Survey of text mining," *Computing Reviews*, vol. 45, no. 9, 2004.

[9] A. McCallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons," in *HLT-NAACL*, 2003.

[10] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*, 2012.

[11] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, 2008.

[12] J. Kleinberg, "Bursty and hierarchical structure in streams," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, 2003.

[13] K. K. Mane and K. Börner, "Mapping topics and topic bursts in pnas," *PNAS*, vol. 101, no. suppl 1, 2004.

[14] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvist. Invest.*, vol. 30, no. 1, pp. 3–26, 2007, publisher: John Benjamins Publishing Company.

[15] D. J. Lafferty, A. McCallum, and C. N. F. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *ICML*, 2001, pp. 282–289.

[16] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and A. N. Smith, "Part-of-speech tagging for twitter: annotation, features, and experiments," in *ACL*, 2011, pp. 42–47.

[17] J. Milgram, M. Cheriet, and R. Sabourin, "One against one or one against all: Which one is better for handwriting recognition with svms," in *IWFHR*, 2006.

[18] R. Socher, A. Perelygin, J. Wu, J. Chuang, D. C. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *EMNLP*, 2013, pp. 1631–1642.

[19] D. Klein and D. C. Manning, "Accurate unlexicalized parsing," in *ACL*, 2003, pp. 423–430.

[20] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, no. 6051, 2011.