

University of Wollongong

Research Online

SMART Infrastructure Facility - Papers

Faculty of Engineering and Information
Sciences

1-1-2016

Social media analysis on evaluating organizational performance: a railway service management

Jie Yang

University of Wollongong, jiey@uow.edu.au

Ahm Mehbub Anwar

University of Wollongong, amanwar@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/smartpapers>



Part of the [Engineering Commons](#), and the [Physical Sciences and Mathematics Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Social media analysis on evaluating organizational performance: a railway service management

Abstract

The last decade has witnessed the dramatic expansion of online social media at the global level. It is widely believed that the influence of social media is of particular importance for public and private organisations. Unfortunately, there is very limited and anecdotal evidence to show the durable influence of social media on various components and functions within organisations. To this end, this study examines the impact of social media, particularly taking the railway services in New South Wales (NSW), Australia and the Twitter platform as an illustrative example. Topic modelling and sentiment analysis techniques are employed to analyse collected tweets for evaluating the organisational performance. More precisely, five topic categories (i.e. reliability, safety, crowdedness, comfort, and convenient) are considered and relevant users' opinions are identified. Statistical results show that discussion on "opal card" is popular when travellers perceive convenient journey, while "peak" and "seat" are the biggest concerns for crowdedness. As a conclusion, the proposed work provides the flexibility and applicability for extracting and understanding complex and huge social-media data, which can in turn be used to inform organisational decision makers.

Keywords

management, social, media, analysis, railway, evaluating, service, organizational, performance:

Disciplines

Engineering | Physical Sciences and Mathematics

Publication Details

Yang, J. & Anwar, A. Mehbub. (2016). Social media analysis on evaluating organizational performance: a railway service management. 2016 IEEE 2nd International Conference on Big Data Intelligence and Computing (DataCom 2016) (pp. 835-841). United States: IEEE.

Social media analysis on evaluating organisational performance: a railway service management context

Jie Yang

*SMART Infrastructure Facility,
Faculty of Engineering and Information Sciences
University of Wollongong, Northfields Avenue, Wollongong,
New South Wales 2522, Australia
Email: jiey@uow.edu.au*

AHM Mehbub Anwar

*SMART Infrastructure Facility,
Faculty of Engineering and Information Sciences
University of Wollongong, Northfields Avenue, Wollongong,
New South Wales 2522, Australia
Email: amanwar@uow.edu.au*

Abstract—The last decade has witnessed the dramatic expansion of online social media at the global level. It is widely believed that the influence of social media is of particular importance for public and private organisations. Unfortunately, there is very limited and anecdotal evidence to show the durable influence of social media on various components and functions within organisations. To this end, this study examines the impact of social media, particularly taking the railway services in New South Wales (NSW), Australia and the Twitter platform as an illustrative example. Topic modelling and sentiment analysis techniques are employed to analyse collected tweets for evaluating the organisational performance. More precisely, five topic categories (i.e. reliability, safety, crowdedness, comfort, and convenient) are considered and relevant users' opinions are identified. Statistical results show that discussion on “opal card” is popular when travellers perceive convenient journey, while “peak” and “seat” are the biggest concerns for crowdedness. As a conclusion, the proposed work provides the flexibility and applicability for extracting and understanding complex and huge social-media data, which can in turn be used to inform organisational decision makers.

1. Introduction

Online social media covers a variety of Internet applications and websites, such as Twitter, Facebook, and Flickr, to name few. More and more people are relying on online social media in their day-to-day lives to access information, express opinions and share experiences with their peers. As a result, massive volumes of online content are generated from numerous social media channels.

A significant proportion of online content is associated with users' opinion. For instance, in the movie domain, social media (such as Rotten Tomatoes) provides cinema fans with convenient mechanisms for posting and sharing their comments about movies. In the e-commerce or e-business area, users may criticize products by leaving online reviews (such as Amazon or Apple store). As a result, investigating online content has attracted considerable in-

terest and curiosity. The benefit of this research work is derived from several aspects. For example, social media analysis can be used to reveal how electronic word-of-mouth (hereafter eWOM) is utilised as a powerful communication tool for spreading awareness of a certain product or service in both the offline and online worlds [1], [2], [3], [4]. Social media is also believed to offer private and governmental organisations a powerful means to improve their knowledge management capability, communication, learning processes and, ultimately, business performance [5], [6], [7]. Consequently, not only can industry stakeholders promote their contents more effectively, but an improved user experience can be offered.

However, regardless of the growing interest in social media analysis in general, there has been little empirical research into the potential benefits and/or commercial applications of social media as a tool for evaluating organisational performance. In addition, very few work focuses on the real-world social media data, that is insufficient to provide a complete understanding of customers' opinion, or measure the effects of both internal and external factors. More importantly, with the exponential accumulation of social media data, the challenges associated with the big *4Vs* problems—data volume (numbers of films, users, and generated reviews), variety (different data formats), velocity (streaming comment data), and veracity (language uncertainty) continue to multiply. This presents a typical scenario for Big Data processing, which is difficult to address using traditional analysis methods.

With this end in view, in this paper we propose a big-data analytical framework to facilitate social-data mining for evaluating organisational performance. Three modules are implemented in this proposed framework, including the data collection, feature generation and machine learning module. The collection module is used to harvest the raw tweet data. Textual features are then generated for different classification purposes. Finally, the quantified features are trained using the Support Vector Machine (SVM) algorithm in the machine learning module. All three modules in the proposed framework are developed based on Apache Hadoop and Spark platforms.

The remainder of this paper is organized as follows. In Section 2, we briefly review some existing work relevant to the impact of social media on organisations. We then introduce the study area including the railway service in New South Wales (NSW), Australia and the Twitter platform in Section 3, followed by three core modules for social-media data mining. The proposed framework is then evaluated using real-world Twitter data in Section 4 and in Section 5 we offer our conclusion.

2. Background

Social media is a special platform that facilitates data sharing and participation among end users [10]. Nowadays, it is increasingly utilised in regular operations within many companies, ranging from start-ups or small enterprises to large corporations. Social media has been regarded as an efficient tool for boosting cooperation and benefits in inter and intra businesses networks. From the organisational view of point, contemporary companies that take advantage of social media seem to outperform their competitors for cost effectiveness and improved efficiencies. Investigators have devised a number of approaches to managerial performance study on social media, as shown in Table 1.

Despite the extensive social media-based analysis, still little is known about the interaction between service provider and end customers. Existing research remains less complete in the social media usage and subsequent impacts on business performance, not to mention the empirical study. Major gaps still exist in current research in the context of the impact of social media on organisational performance.

3. Methodology

In this section, we first provide some background information on the study area, including the popular social media platform (Twitter) and the railway service in NSW, Australia. Next, a big-data processing framework is proposed that consists of three main modules. Employed big-data techniques and classification algorithm are also elaborated herein.

3.1. Study area

This study has been designed by focusing on the railway services in NSW using the social media data. NSW Government follows several social media platforms to improve their services as per customers' demand. There are mainly three platforms: online websites (such as TfNSW [18]), Facebook, and Twitter. Rail organisations are very much familiar of interacting with customers via these social networks to updating traffic information. For instance, on April 2015, NSW was affected by a very strong storm. During this period, these social platforms had recorded numerous visitor activities:

- TransportNSW.info had 250,000 users who turned online for help over a 48 hour period (nearly double amount of visitors);

- Live Traffic NSW website and mobile apps had more than 482,000 visitors (about 20 times more than usual).
- 2.2 million people visited Live Traffic NSW Facebook page within 7 days;
- Tweets (on 22 April, 2015), about the flooding at Bardwell Park, were seen about 174,000 times;
- 375,000 tweets were found related to the storm damage.

Among three social platforms, Twitter is an public and real-time information service, which records daily activities of people. On a daily basis, millions of messages, pictures, URL links are posted in this platform, that results in a dynamic data streaming. In this paper, we are taking Twitter as the main social media resource.

3.2. Proposed framework

In this section, we present a big-data processing framework as a means of discovering critical patterns towards customers' opinion. The proposed framework consists of three main modules, a brief description of which is given as follows:

- *Data collection module*: raw tweet data are collected and distributed;
- *Feature generation module*: this module helps in identifying significant attributes from the raw tweets and producing high-level features;
- *Machine learning module*: the SVM algorithm is applied in this module to build up the classification model for topic modelling and sentiment analysis, repetitively.

3.2.1. Data collection module. Twitter allows access to its data via public Application Programming Interfaces (APIs). In this paper, we collected live tweets using the Twitter streaming API inside NSW, Australia from March to November, 2015.

Furthermore, to pull out useful tweets from streaming data, some keywords are employed to eliminate irrelevant tweets. Those keywords (as shown in Table 2) are selected based on names of places or organisations (such as "train", "rail", "131500", "station"), or general terms for travelling (such as "travel", "safe", and "comfort"). In practise, regular expressions is applied to the streaming data, and we only keep the tweets that contain defined keywords.

TABLE 2. KEYWORDS USED FOR RAW TWEET COLLECTION

train	rail	131500	station	commute
fare	transport	travel	safe	comfort

In this collection, two categories of tweet contents are considered: tweet content and user profile. (The raw attributes and relationships between contents are represented

TABLE 1. SUMMARY OF STUDIES RELATED TO SOCIAL MEDIA AND ORGANISATIONAL PERFORMANCE

Issues/dimensions	Findings	Sources
social media and organisations	Individual’s competence varies on the shape and size of his/her social networks; and the competence plays an important role to manage organisation efficiently and become more profitable.	[8]
knowledge management in supply chains	Social media establishes a relationship among the companies along with supply chain, exchanges the knowledge to create specific capabilities in organisational management.	[9]
CEO’s social media management and organisational performance	CEOs’ social capability influences the organisational performance through knowledge and strategic flexibility.	[10]
Knowledge conversion and team performance	Interaction through social media among team members and intra-organisational networks contributes to team performance and ultimately organisational performance.	[11]
Social capital, knowledge transfer and performance	Knowledge transfer acts as an instrument for developing social capital.	[12]
Knowledge management in small and medium enterprises	Social capital and motivation-opportunity-ability models are useful to investigate human factors to characterise knowledge sharing from both social and technological dimensions perspective.	[13]
Networks in firm performance	Strong and heterogeneous networks improve innovation towards performance.	[14]
Organisational knowledge sharing	Social media motivates a person to exchange knowledge and opinions.	[15]
Knowledge sharing and firm innovation capability	Individual factors, such as enjoyment in helping others using social media, affect knowledge-sharing processes significantly, thereby improving innovation capability of the firms.	[16]
Strategic flexibility	Social media affects strategic flexibility positively.	[17]

in Figure 1). Collected samples are then distributed and stored separately according to the data category. That is, for any individual day, tweets are recorded in one log file, while the related user profile is kept in another log document. One record (either one tweet or a user information) is represented as a single row in the log file, while the special symbol “#” is used to separate raw attributes. Again one log file is used for records from the same day only.

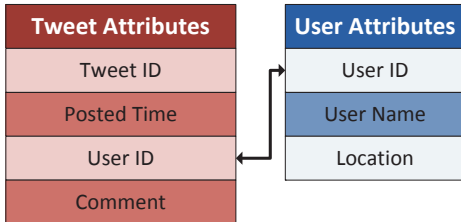


Figure 1. Data structure and relationships between tweet contents and users.

3.2.2. Feature generation module. To interpret the collected raw tweets, that is, to identify and extract subjective information from the source material and turn it into usable data, the feature generation module is implemented to facilitate the data analysis. A feature can be regarded as a user-defined hierarchical representation for the initial content (aka the raw tweets). Features are then used later for the classification purpose in the machine learning module.

Before generating any textual features, the preprocessing procedure is firstly employed to cleanse the raw tweets. The preprocessing aims to eliminate unnecessary contents from raw data, including: (1) isolated @ signs; (2) user names; (3) URL links; (4) punctuations (such as '&', '!', '*', '#', '\$'); (5) English stopwords, like “the”, “a”, “and”, *etc.* Next,

TABLE 3. SUMMARY FOR GENERATED TEXTUAL FEATURES REPRESENTING RAW TWEETS.

Feature	Description
TF-IDF	term frequency (TF) represent the number of times that a word occurs in the document; and inverse document frequency (IDF) is the inverse document frequency of a term.
POS tags	part-of-speech tags.
Emoticons	icons showing user’s emotion, such as a smile symbol.
Tweet length	how many characters within one tweet.
Term presence	number of positive or negative words.

textual features are generated using all cleansed tweets (as summarized in Table 3).

Among all features, the TF-IDF feature reflects how important a word is to a document in a corpus. The POS tag is the identification of words as nouns, adjectives, *etc.* For this POS feature, we are only selecting terms that are nouns, adjectives, adverbs and verbs. Such a selection is made based on their relevance in determining sentiment, as they are more associated with users’ opinion than other terms. Therefore, terms with other POS tags are eliminated in this feature. Herein we consider the number of POS tags occurred in a sample tweet. As an example, one collected tweet was : “problems bike availability bike share n’t one yet safety cmprd os cycling syd = dangerous”. In particular, a sequence of 9, 2, 0, 1 terms is matched to the noun, adjective, adverb and verb tag, respectively. Thus, the related POS tags feature is represented as: (9, 2, 0, 1) for this tweet.

Emoticon is very commonly used in the Twitter platform, such as a smile symbol or sad face. In this paper, pos-

itive emoticons (such as a smiling face) are given a weight of “1” and negative emoticons (like angry or sad symbol) are given a weight of “-1”. In those tweets without any emoticons, we give them a zero value. At last, occurrence of positive and negative terms in tweets are also calculated as two different features. The polarity corpus in SentiWordNe is introduced to identify positive and negative terms [19].

3.2.3. Machine learning module. To identify and extract valuable information from raw tweets, the machine learning module is implemented for two main purposes: topic modelling (or document classification) and sentiment analysis. In general, topic modelling involves the division of tweets into their semantic topics. In this paper, the topics are predefined as “safety”, “reliability”, “crowdedness”, “convenient”, and “comfort” because these topics are the keys to measure the performance of a railway service. To express the matter in mathematical terms: let $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ be a tuple representing n -textual features extracted from one tweet, and $y \in \{safety, reliability, crowdedness, convenient, comfort\}$ be the class label. Topic modelling aims to train a classifier that extracts the decision rule subject to the following constraint:

$$y = f(\mathbf{x}) + e, \quad (1)$$

where $f(\cdot)$ is an decision function to be estimated by the classifier, and e is the corresponding error.

For sentiment analysis, we would like to categorize tweets as positive, neutral or negative. Similar to topic modelling, let $z \in \{positive, neutral, negative\}$ be the class label. Sentiment analysis aims to train a classifier that satisfies the following constraint:

$$z = g(\mathbf{x}) + e, \quad (2)$$

where $g(\cdot)$ is the decision function.

The Hadoop framework is employed in this paper as the fundamental component to save large amount of tweets, whilst the machine learning module is implemented on top. The Hadoop framework is commonly used to analyse big data sets, such as social network, online web content, or large-scale graph [20]. In addition, the machine learning module is implemented using the Spark framework. Spark is an open source environment for fast data analytic [21]. It provides a scalable platform for in-memory computing, thereby achieving advanced performance over other approaches. MLlib is a machine learning library running under the Spark platform, in which we build up our own classifiers. The final outcome is again stored back into Hadoop. Fig. 2 illustrates the implemented ecosystem for the machine learning module.

In addition, the *Support Vector Machine* (SVM) algorithm is implemented as the classifier. As a supervised learning approach, the SVM algorithm has been demonstrated to perform well in various practical applications [22]. Furthermore, for multi-class problems, the one-against-one strategy is implemented [23]. Therefore, $\frac{N_{class} \times (N_{class} - 1)}{2}$ SVM classifiers is constructed if there are N_{class} classes.

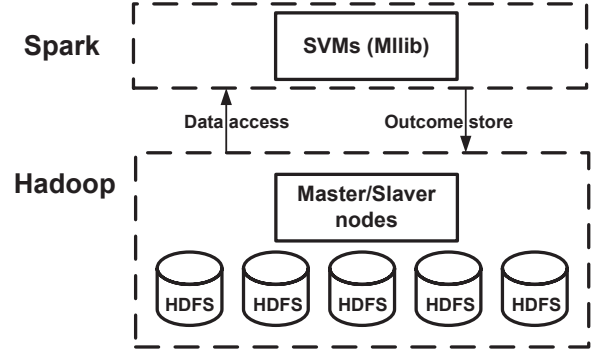


Figure 2. Implemented ecosystem for the machine learning module.

4. Data analysis

In this section, we analyse the Twitter data to characterize their contribution in the organisational service context. The employed cloud infrastructure is presented in Subsection 4.1. The experimental setup and data sets are presented in Section 4.2. The analytical outcome is then presented in Section 4.3.

4.1. Cloud infrastructure

Virtual cluster is a simple but fast environment to build up the big data ecosystem, i.e., Hadoop and Spark framework. In our implemented cloud infrastructure, a Dell server with Intel Xeon E5-2630 1.8GHz cores and 32G memory is employed. A virtual cluster of four nodes is then deployed. For each node, two virtual CPU and 4GB of memory is allocated. In addition, one node is set up as the master machine for both Hadoop and Spark, while the rest is used as the slaver node. In addition, for the Hadoop platform, the 1.2.1 version is installed. Accordingly, we take Spark 1.5.0 as the running version and the standalone model is adapted to cope with the Hadoop framework.

4.2. Experimental setup

Table 4 shows the summary statistics for the data harvested from Twitter from March until November 2015. Again we only collect tweets containing keywords in Table 2, that lead to a total of 31,008 tweets from 9,428 distinct users. For each single day, on average, around 115 tweets are collected.

TABLE 4. DESCRIPTIVE STATISTICS RELATING TO COLLECTED TWEET DATA

Variable	Mean	Standard deviation	Total
Number of tweets	114.8	25.9	31,008
Number of users	34.9	10.7	9,428

Figure 3 illustrates the geographic distribution of users who participated in the study. About 81.39% of users left location information. Unsurprisingly, most users were located

in big cities such as Sydney, Melbourne, Canberra, while users from UK formed the largest single overseas group. Meanwhile, there is another two large group of users from Australia (21.91%) and New South Wales (12.27%), respectively, although they did not leave further detail. Generally speaking, in these places users have a convenient access to social network and are more willing to spare their opinions. Therefore, the data illustrates significant participation from these areas.

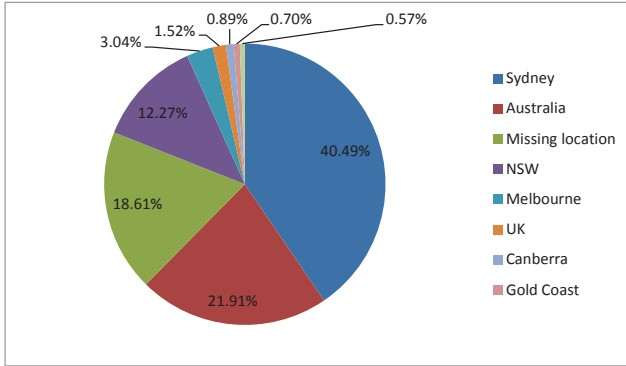


Figure 3. Geographic distribution of twitter users in the study.

For classification, the training parameters for the SVM algorithm are set as follows: the maximum number of training iterations is 200; the default L2 regularization with the regularization parameter 0.05 is performed during the iteration; the updating step size is 1.0; and the minimal batch fraction is set as 0.8. The generalization performance is evaluated using the classification error.

For both topic modelling and sentiment analysis, we first randomly select 30% tweets from the entire data set and then manually label them with both topic categories and sentiment. Then these samples are used for training two multi-class SVM classifiers for topic modelling and sentiment analysis, respectively. Finally, we apply the trained SVM classifiers to identify the remaining unlabelled tweets.

4.3. Performance analysis

4.3.1. Topic modelling. Topic modelling aims to find for commonalities in a series of texts and group them into predetermined labels based on topical themes. In this paper, reliability, safety, crowdedness, comfort and convenient are selected as target topic themes. By classifying tweets according to different topics, we can then find out the most concerned or popular word-of-mouth from customers.

Figure 4 illustrates the ROC curve for topic modelling while categorizing tweets for different topics. As observed, the trained SVM achieved average 90.47% classification accuracy for topic modelling. We then apply this classifier to category the unlabelled tweets to find out their topics.

Table 5 presents the statistics of the distribution about tweet topics. It is clearly seen that people talk most about the comfort of service, while reliability is another important

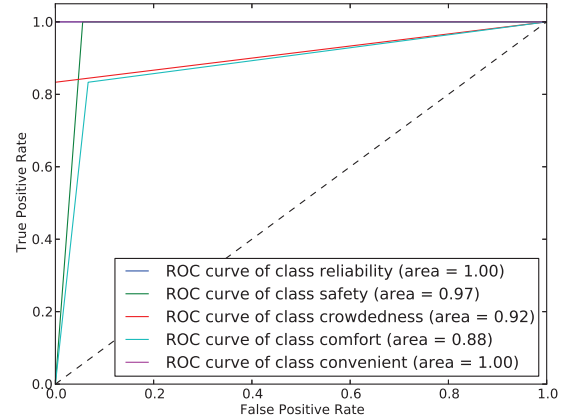


Figure 4. ROC curve for the classification accuracy of topic modelling.

topic for train customers. By contrast, crowdedness of train is less discussed topic in this study.

TABLE 5. STATISTICS OF CLASSIFIED TWEETS IN TERMS OF TOPICS AND USER BACKGROUND

Topic	Percentage
reliability	22.8%
safety	15.8%
crowdedness	6.9%
comfort	36.6%
convenient	17.9%

TABLE 6. MOST FREQUENT TF-IDF KEYWORDS FROM CLASSIFIED TWEETS

Topic	Top 5 TF-IDF words
reliability	(sacrifice, 0.241), (sucks, 0.241), (waiting, 0.120), (missed, 0.120), (slow, 0.120)
safety	(safe, 0.333), (safety, 0.333), (lost, 0.166), (bike, 0.166), (station, 0.166)
crowdedness	(peak, 0.420), (cuts, 0.210), (lift, 0.210), (seats, 0.110), (lots, 0.020)
comfort	(ok, 0.150), (sweat, 0.150), (uncomfortable, 0.150), (stop, 0.150), (full, 0.150)
convenient	(opal, 0.372), (card, 0.206), (machine, 0.186), (station, 0.140), (money, 0.070)

Table 6 further demonstrates the most informative TF-IDF keywords that related to users' opinions. For reliability, identified keywords are related to "waiting" time and "slow" service. Meanwhile, people also complain some "sacrifice" due to the "missed" train. Therefore, regular service and minimal waiting time are expected to enhance the reliability performance. In terms of safety, the keyword "bike" is surprisingly identified. In fact, in some raw tweets people are criticising the cycling condition inside the Sydney CBD area. More public transportation service is expected to offer a better solution to resolve the traffic congestion. As for the crowdedness issue, the word "peak" time is discussed prominently (42%) due to the increased people flow on the train. Thus, flexible (or adjusted) departure time may

be anticipated to minimise the crowdedness at peak hours. Furthermore, “lift” availability carries approximately 21% importance, which indicates the lift facility inside the train station should also be taken into consideration by the service provider to improve the crowdedness condition. Last, the opal card, an electronic ticketing service in Sydney Train, is the highly (37%) reported keyword in the convenient topic.

4.3.2. Sentiment analysis. Sentiment analysis in this paper is used to understand customers’ satisfaction towards target topics, i.e., reliability, safety, crowdedness, comfort and convenient. Figure 5 illustrates the ROC curve for sentiment analysis. As observed, we achieved average 91.32% accuracy for three classes. We then apply the trained classifier to find out customers’ opinion for the rest tweets.

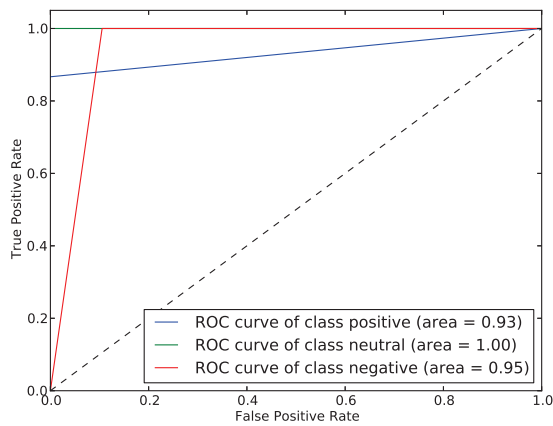


Figure 5. ROC cure for the classification accuracy of sentiment analysis.

Figure 6 illustrates the distribution of users’ opinions for five topics. As observed, negative sentiment is always at a higher level compared to positive and neutral sentiments. That is, service users are dissatisfied with the provided rail service as most of tweets are showing negative opinion on the Twitter platform. In particular, in terms of the crowdedness, people are highly dissatisfied (0.92) with the services because they have congestion experiences in peak hours. The reliability performance has also been mostly negatively expressed (0.68), which indicates users’ unhappiness for the long waiting time.

By contrast, users’ experience with the rail convenience has a positive outcome, as their sentiments turn out to be very close (positive: 0.34; neutral: 0.31; negative: 0.34). The possible reason could be the opal card service is getting more popular, and people are happy with this electronic ticketing system.

5. Conclusion

An important part of information-gathering behaviour has always been to find out what people think. With the growing availability and popularity of opinion-rich resources, such as Twitter, new opportunities and challenges

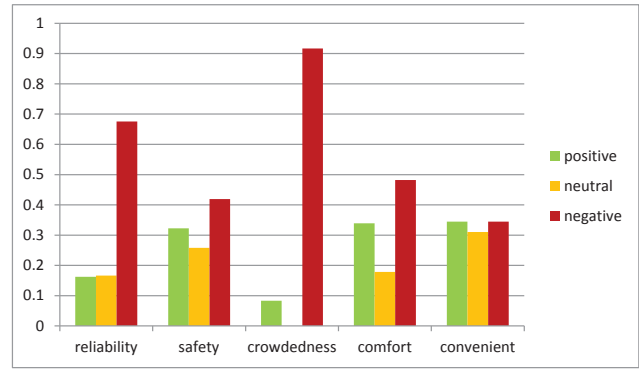


Figure 6. Distribution of users’ opinions for five topics.

arise. People now can, and will, actively use information technologies to seek out and share the opinions with others. The eruption of activity in social media has thus attracted great interest and new development.

Our focus in this paper is on methods that seek to address the new challenges raised by train users for the rail service improvement using the Twitter data. A big data analytic framework is proposed in this study by implementing three main modules: data collection, feature generation, and machine learning (for topic modelling and sentiment analysis), thereby providing a new understanding about the potential improvement pathway of the service. For example, by looking at most frequent topics or keywords, rail organisations can make a better decision for the infrastructural planning. Meanwhile, by conducting the sentiment analysis on collected tweets, we will have a better understating of users’ satisfaction, which can contribute to a better users’ experience.

In conclusion, the proposed analysis reveals interesting patterns related to the rail service performance. It can further help maximizing users’ satisfaction towards the service and assisting in the formulation of a policy response.

References

- [1] V. Schnberger and K. Cukier, “Big data: A revolution that will transform how we live, work and think.” in *New York: Houghton Mifflin Harcourt Publishing Company*, 2013.
- [2] J. Yang and B. Yecies, “Mining chinese social media ugc: a big-data framework for analyzing douban movie reviews,” *Journal of Big Data*, vol. 3, pp. 1–23, 2016.
- [3] J. Yeap, J. Ignatius, and T. Ramayah, “Determining consumers most preferred ewom platform for movie reviews: A fuzzy analytic hierarchy process approach,” *Computers in Human Behavior*, vol. 31, pp. 250 – 258, 2014.
- [4] H. Hu, Y. Wen, Y. Gao, and T. Chua, “Toward an sdn-enabled big data platform for social tv analytics,” *IEEE Network*, vol. 29, pp. 43–49, 2015.
- [5] S. Andriole, “Business impact of web 2.0 technologies,” *Communications of the ACM*, vol. 53, no. 12, pp. 67–79, 2010.
- [6] J. Twentyman, *Business starts to take Web 2.0 tools seriously*. Financial Times, London, 2009.

- [7] Z. Liang, H. Zou, J. Guo, and F. Yang, "Multi-user web service selection based on multi-qos prediction," *Information System Frontiers*, vol. 16, pp. 143–152, 2014.
- [8] V. Zyl, "The impact of social networking 2.0 on organisations," *The Electronic Library*, vol. 27, no. 6, pp. 906–918, 2008.
- [9] J. Capo, J. Mula, and J. Capo, "A social network-based organizational model for improving knowledge management in supply chains," *Supply Chain management: An International Journal*, vol. 16, no. 5, pp. 379–388, 2011.
- [10] V. Perez, V. Morales, and F. Sanchez, "The effects of ceos social networks on organizational performance through knowledge and strategic flexibility," *Personnel Review*, vol. 41, no. 6, pp. 777–812, 2012.
- [11] M. Janhonen and E. Johanson, "Role of knowledge conversion and social networks in team performance," *International Journal of Information Management*, vol. 31, no. 3, pp. 217–225, 2011.
- [12] I. Maurer, V. Bartsch, and M. Ebers, "The value of intra-organizational social capital: How it fosters knowledge transfer, innovation performance, and growth," *Organization Studies*, vol. 32, no. 2, pp. 157–185, 2011.
- [13] M. Chen, C.W. Chang and C. Tseng, "Human factors of knowledge-sharing intention among taiwanese enterprises: A model of hypotheses," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 22, p. 362371, 2012.
- [14] S. Gronum, M. Verreyne, and T. Kastelle, "The role of networks in small and medium-sized enterprise innovation and firm performance," *Journal of Small Business Management*, vol. 50, no. 2, pp. 257–282, 2012.
- [15] S. Chow and S. Chan, "Social network, social trust and shared goals in organizational knowledge sharing," *Information & Management*, vol. 45, no. 7, pp. 458–465, 2008.
- [16] F. Lin, "Knowledge sharing and firm innovation capability: An empirical study," *International Journal of Manpower*, vol. 28, no. 3, pp. 315–332, 2007.
- [17] V. Perez, M. Fuentes, and A. Bojica, "Strategic flexibility and change: The impact of social networks," *Journal of Management and Organization*, vol. 18, no. 1, pp. 2–15, 2012.
- [18] [Online]. Available: <http://www.transport.nsw.gov.au>
- [19] [Online]. Available: <http://http://sentiwordnet.isti.cnr.it/>
- [20] [Online]. Available: <http://hadoop.apache.org/>
- [21] [Online]. Available: <http://spark.apache.org/>
- [22] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.
- [23] J. Milgram, M. Cheriet, and S. Robert, "One against one or one against all: Which one is better for handwriting recognition with SVMs," *Tenth International Workshop on Frontiers in Handwriting Recognition*, 2006.