Scientometrics (2021) 126:6383–6421
https://doi.org/10.1007/s11192-021-04046-2

Investigating transportation research based on social media analysis: a systematic mapping review

Tasnim M. A. Zayet[1] · Maizatul Akmar Ismail[1] · Kasturi Dewi Varathan[1] · Rafidah M. D. Noor[2] · Hui Na Chua[3] · Angela Lee[3] · Yeh Ching Low[3] · Sheena Kaur Jaswant Singh[4]

Received: 27 December 2019 / Accepted: 17 May 2021 / Published online: 24 June 2021

Abstract
Social media is a pool of users' thoughts, opinions, surrounding environment, situation and others. This pool can be used as a real-time and feedback data source for many domains such as transportation. It can be used to get instant feedback from commuters; their opinions toward the transportation network and their complaints, in addition to the traffic situation, road conditions, events detection and many others. The problem is in how to utilize social media data to achieve one or more of these targets. A systematic review was conducted in the field of transportation-related research based on social media analysis (TRR-SMA) from the years between 2008 and 2018; 74 papers were identified from an initial set of 703 papers extracted from 4 digital libraries. This review will structure the field and give an overview based on the following grounds: activity, keywords, approaches, social media data and platforms and focus of the researches. It will show the trend in the research subjects by countries, in addition to the activity trends, platforms usage trend and others. Further analysis of the most employed approach (Lexicons) and data (text) will be also shown. Finally, challenges and future works are drawn and proposed.

Keywords Intelligent transportation system · Traffic · Opinion mining · Text mining · Social media analysis · Systematic mapping review · Sentiment analysis

✉ Maizatul Akmar Ismail
maizatul@um.edu.my

[1] Department of Information Systems, University of Malaya, Kuala Lumpur, Malaysia

[2] Department of Computer System and Technology, University of Malaya, Kuala Lumpur, Malaysia

[3] Department of Computing and Information Systems, Sunway University, Petaling Jaya, Selangor, Malaysia

[4] Department of English Language, Faculty of Languages and Linguistics, University of Malaya, Kuala Lumpur, Malaysia

Introduction

Social media platforms have rapidly gained its user base since its emergence in the late 1990s. Since then, people have started to publish their thoughts, opinions, experiences, ratings and even their daily lives' details through social media applications such as Facebook, Twitter, Weibo and many others. Consequently, social media data has become a great source of real-time data that can be useful for many sectors. In politics, social media data is used to predict election results (Caetano et al., 2018; Jaidka et al., 2019); in e-commerce, it is used to create recommendations; in economy, it is used to make inferences about the stock market movements (Bollen et al., 2011; X. Zhang, Zhang, et al., 2018; Zhang, Chen, et al., 2018); in tourism, it is used to recommend the best hotels, experiences and places (A. S. H. Lee et al., 2018) and in education, it is used for teaching and learning resources (Pereira et al., 2018).

The real-time property of social media data has been debated in numerous studies. These studies found that social media is a source of real-time or near-real-time data, especially in the case of accidents and crises, because people post on their accounts immediately after an incident occurs (Daly et al., 2013; Osborne et al., 2014; Y. Gu et al., 2016; S. Zhao et al., 2017). So, in time-sensitive domains such as transportation, real-time social media data can be very useful; timing has the potential to influence users' everyday activities and even save lives. In the transportation network, any delays could result in severe problems, for example, people being late for their work and schools, ambulance delays, or crowdedness of stations and others. Usually, transportation companies use physical sensors (Guerrero-Ibáñez et al., 2018) to collect information about traffic situations and delays. These sensors are costly and need maintenance. In addition, any defect or attack on the sensors network can cause problems (X. Zhang et al., 2017) in the delivery of information, which can lead to problems in the traffic system and organization.

In case of defects or faults in the sensors' network, other real-time data sources are needed for troubleshooting. One of the most suitable sources, in the aspects of time and budgeting, is social media data. Social media data can provide real-time information about various transportation statuses from the commuters' perspective. It can help in extracting road hazards (Kumar et al., 2014), traffic situations (D'Andrea et al., 2015; D. Wang et al., 2017), event or accident detection (Candelieri & Archetti, 2015) and others.

Furthermore, people in charge of the transportation network can have inferences based on commuters' opinions and complaints (Abalı et al., 2018) toward transportation networks (Adeborna & Siau, 2014; Gal-Tzur et al., 2014). They need these inferences to form opinions on the best paths to take in order to maximise network utilisation. Commonly, traditional data collection approaches such as questionnaires and surveys are used by decision-makers to gain an understanding of public opinion (Pournarakis et al., 2017). Those traditional methods are time-consuming and costly in contrast to social media data use.

Due to the reasons above, employing social media data in transportation-related studies has become a trend. Social media data can be obtained by performing a simple crawling of the network. Therefore, this paper presents a systematic review of transportation studies that have adopted social media analysis (TRR-SMA) and have been published between the years 2008 and 2018. It is showing the structure of the field by providing information on targeted topics, types of data used, data collection methods and data analysis methods; thus, this paper provides the new researchers with comprehensive information on how the social media data in transportation researches had been utilized in the past and for which goals. In addition, it also provides in depth information on the challenges, issues and

research gaps that future researchers should embark; hence, it can serve as a starting exploration point for new TRR-SMA researchers.

As for the rest of the paper, it is organized as follows: Firstly, an overview of the existing survey papers is presented. Secondly is the presentation of the research methodology of this study. Thirdly, the report of our systematic mapping review analysis is provided. Fourthly, the discussion of the main findings and possible future works are laid out. Finally, the conclusion of the paper is presented. In this study, terms such as papers, literature, works and studies are of the same meaning. In addition, applications and platforms are also considered as synonyms.

Prior work

There are several review papers in the field. In producing review papers, three key factors affect the reported papers: the query, the time of the publications and the digital libraries used in the searching process. Different review papers can be generated by changing at least one of any of those factors. Each existing review paper varies from our paper in different aspects: the query used in the searching process, the period of the published literature, reported data and its type, and the methodology of the analysis. In the following work, we present each review paper, then subsequently, our systematic mapping review analysis.

Nikolaidou et al. (Nikolaidou & Papaioannou, 2018) presented a review paper of used social media data, methods and challenges in transport planning and public transit quality. Nikolaidou et al. reviewed around 50 papers which were published between 2010 and 2016. Lv et al. (2017) proposed a literature review of social media-based transportation research using social network analysis method and 67 papers for the period between 2011 and 2015. They showed the research collaboration in the field based on the authors, institutions and countries. Rashidi et al. (2017) proposed a review of social media data used for modelling travel behaviour with its advantages and challenges. They had ended with 800 + papers from Scopus for analysis which were published between 2007 and 2015 but they had mentioned around 70 papers in their review paper. However, as they were aiming on modelling the travel behaviour, they focused on the location data. Chaniotakis et al. (2016) produced a mapping review of studies that used social media data and platforms. In addition, the authors addressed the research challenges and opportunities in using social media for transportation studies. Chaniotakis et al. review was based on 22 literature published between 2008 and 2015. Grant-Muller et al. (2014) proposed a review using around 70 papers published between 2008 and 2014. The review covered the social media data that could be used in transportation-related research. They performed a more in-depth study and argued whether social media data could be used along with the transportation data. In addition, they presented text mining methods and the challenges in the transportation field.

Comparatively, the analysis method of our literature review was conducted in a more systematic way. The selection process of our primary studies was designed to ensure reliability and reproducibility. This implies that our results of analysis can be regenerated by following the steps presented in the Methodology section. In most of the above-mentioned review papers, the used query was not reported, so placing a comparison based on query was not possible. However, (Rashidi et al., 2017) was the only one to report the used query which is different from the one we used. We shape the query using general terms since we are looking at the field from a broad perspective to extract the used social media platforms, data, and data analysis approaches, as well as the research targets. In the broadest sense, it is important to provide a large scale and coverage of the data being analysed. Limiting any

Fig. 1 Systematic mapping construction methodology

of the targeted data to predefined types and categories would reduce the number of outcomes, resulting in the omission of some data forms. So, we did not use any name of social media platforms in forming our query.

In addition, our paper presents a review of the literature published in the years between 2008 and 2018 from four digital libraries, suggesting that the review has a broader scope than previous studies. From the aspect of the analysed data classification, we present a more fine-grained examination by analysing various components while others focused on one of these components. The components that have been included in the analysis are the distribution of the publication over countries, first authors, publishers and years, keywords analysis, focusses of transportation-related research and types of the used social media data and platforms. To the best of our knowledge, this paper is the first to analyse the research subjects by countries and years in order to show the trends in the relevant research field, as well as the first to present the trend analysis of social media platforms used for TRR-SMA. Furthermore, it is the first to analyse the text data attributes according to the aims of using them.

Methodology

We adopted a systematic mapping review methodology to provide an overview and research structure of the field of TRR-SMA (Petersen et al., 2008). This paper differs from systematic literature reviews (SLRs) as the latter aims to identify, analyse, evaluate and report the existing research in a field using well pre-defined and repeatable steps. These steps generate the primary study set (Kitchenham et al., 2009). The primary set contains the papers resulted from the search and selection process that will be reported in the paper. Meanwhile, a systematic mapping study or scoping study aims to structure and give an overview of the field of interest by classifying the papers in the primary study set and analysing them according to their numbers in the categories (Petersen et al., 2008).

In conducting our study, we adapted the guidelines provided by (Kitchenham et al., 2009) and (Petersen et al., 2008) with slight modifications to suit the context of our research objectives. Kitchenham and Charters proposed the guidelines for generating an SLR in the field of software engineering, while Petersen et al. presented guidelines of a systematic mapping method in the same field. A systematic mapping study for the software engineering field, based on a similar methodology, was found in another study (Zakari et al., 2018).

The main steps suggested by (Petersen et al., 2008) and adopted in this study are shown in Fig. 1. The first step is defining the research questions (RQs) that will be answered by the study. The second step is constructing the search protocol and search, while the third step is screening the search results using the inclusion and exclusion criteria. The fourth step is constructing the classification scheme and defining the categories, and finally, extracting the desired data from the primary set and mapping it to the categories. The following sub-sections describe each step of our research process in detail.

Table 1 Main research questions (RQs)

Research questions (RQs)	Motivation
RQ1: How social media is used in transportation research based on social media analysis?	To identify the trends in the field, the used keywords, the used social media data, the used social media platforms and the used approaches
RQ2: What are the aims of transportation researches based on social media analysis?	To identify the targets of the researches and their trends in the world and by countries
RQ3: What are the challenges, principal findings and possible future works in the field?	To identify the challenges in the field and main findings from the analysis and draw the needs of the field

Table 2 Sub-questions of RQ1

Sub-research questions drawn from RQ1 (s-RQ1)	Motivation
s1-RQ1: What is the distribution of the researches in terms of activity?	To identify the trend in the publication in the field in terms of years, countries, publishers and first authorship
s2-RQ1: What are the used keywords in the field?	To identify the most used keywords by authors in the researches
s3-RQ1: What are the social media data/attributes used by the researchers?	To identify the used social media attributes, the most used ones and the aim of using them
s4-RQ1: What are the rules of text data/text mining in the TRR-SMA field?	To identify the usages of the text data
S5-RQ1: What are the social media platforms used by the literature?	To identify the most used social media platforms and their usage trend
S6-RQ1: What are the datasets used by researchers?	To identify the datasets and the methods of collecting them
S7-RQ1: What are the approaches used to analyse social media data in transportation researches based on social media analysis?	To identify the most used methods for analysing social media data

Defining the research questions

The ultimate aim of this study is to investigate the researches in the field of social media-based transportation. By investigating the researches, we find the elements/foundations that have to be taken into consideration when starting any research in the field, hence, it directs the way of utilizing social media in transportation-related researches.

Three main RQs have been identified. Table 1 shows these RQs with their respective motivations. The first and second RQs lead to multiple sub-RQs (s-RQs) as they implied different data to be analysed, hence, these s-RQs are forming the subsections to present the results in a more organized way. The s-RQs are shown in Table 2 and Table 3. As for the RQ3, it poses the challenges, findings and future works; since these three are inextricably linked and lead to one another, they will be addressed for each analysed data rather than being divided into s-RQs.

Table 3 Sub-questions of RQ2

Sub-research questions drawn from RQ2 (s-RQ2)	Motivation
s1-RQ2: Which subjects were targeted by researchers in the TRR-SMA field?	To identify the targets of the researches
s2-RQ2: What are the trended subjects in the world and by countries?	To identify the trends of research-subjects around the world and in the target countries
s3-RQ2: What are the social media attributes used for achieving the targets?	To identify the role of social media data in the research field to achieve each research target

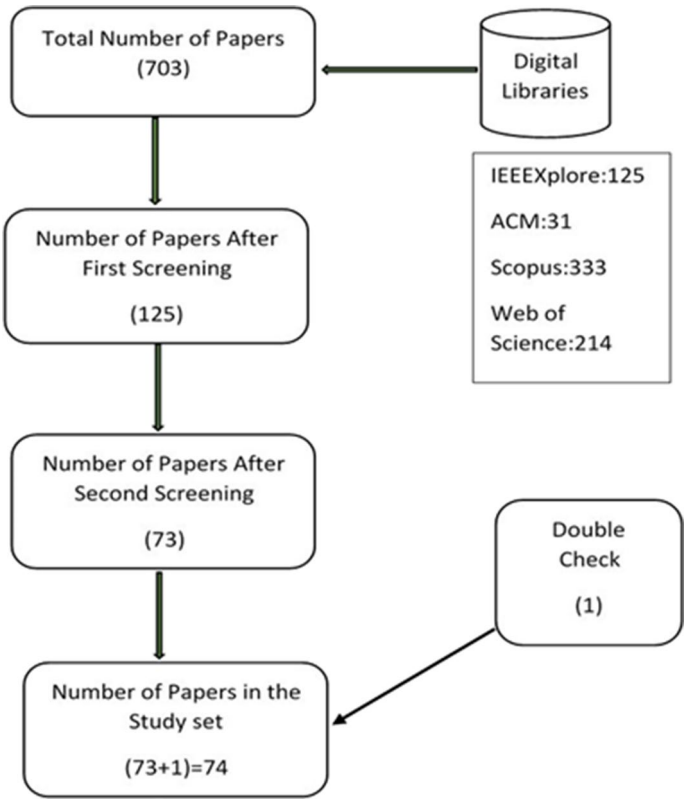


Fig. 2 Search and selection process

Search and selection process

The search protocol contains three steps: Defining the digital sources, constructing the search query and lastly, applying the search process. Figure 2 shows an overview of the search and selection process.

Table 4 Query terms and synonyms

Transportation-related terms	Social media analysis-related terms
Transport	Social media analysis
Transportation	Opinion mining
Transport-related	Text mining
Intelligent transportation system	Sentiment analysis
	Social network analysis

Defining digital sources

Four digital libraries (DLs) were chosen to refine the selection of papers: IEEE Xplore, ACM, Web of Science and Scopus. Google Scholar was not considered as it contains too large a proportion of irrelevant literature to this study, in addition to grey literature.

Forming the search query

Based on the defined research questions in Section "Defining the research questions", terms and keywords were identified. These terms and keywords were revised according to the pre-scanned literature. Furthermore, we revised them by including the synonyms. The terms and keywords are illustrated in Table 4. By combining the mentioned terms, the following query was constructed:

(transport* AND ("sentiment analysis" OR "opinion mining" OR "text mining" OR "social media analysis" OR "social network analysis"))

General terms were used in our query. Instead of using the names of social media platforms that we hope to learn from the studies, we used the term "social media analysis" and its synonyms to describe the process of analysing social media data.

Conducting the search

The search was performed on the 16th of January 2019 using the previous mentioned DLs and by searching through titles, abstracts and keywords. The total number of search results was 703 as illustrated in Fig. 2.

Defining the inclusion and exclusion criteria

Any paper to be included in the primary study set has to fulfil the inclusion and exclusion criteria to assure its relevance to the field and its ability to answer the RQs. In other words, the paper has to fulfil all the inclusion terms and none of the exclusion terms. The defined Inclusion (IC) and Exclusion (EC) terms are shown in Table 5. In case of IC4 and EC4, the extended version was included because, usually, it provides more information about the procedures, the experiment, the findings, and the assessment, as opposed to reporting the process without results or evaluation or reporting the results

Table 5 The inclusion and exclusion criteria

Inclusion terms (ICs)	Exclusion terms (ECs)
IC1: The work proposed a method for social media analysis for transportation-related subjects	EC1: The work is a thesis, book and other grey literature
IC2: The work is a journal paper or a paper in a conference proceeding/ peer-reviewed paper	EC2: Papers written in other languages, other than English
IC3: Clearly mentioned the dataset used	EC3: The dataset is not clear/mentioned
IC4: Workshop/journal papers that have been extended from conference papers	EC4: Conference papers that have been extended to journal/workshop papers
	EC5: Works related to social media analysis but not for transportation-related subject or vice versa

Table 6 Examples of the excluded studies and the corresponding EC term

The excluded studies	Exclusion term (EC)
Liu et al. (2018)	EC2
Patel et al. (2013)	EC3
Di Wang et al. (2014)	EC4

based on only a portion of the data. Moreover, worth to mention that in our exclusion and inclusion criteria, the citations do not play any rule.

The search results were refined in two screening stages using the IC/EC terms, the first stage, according to the title and abstract and the second stage, according to the full text. Table 6 shows examples of excluded studies and the corresponding exclusion term.

Conducting the classification scheme

The construction of the classification scheme depends on the prior knowledge of the field and the first screening phase of the literature. The prior knowledge was gained by scanning existing reviews and literature in the field. The first screening was applied using the abstract and the title of the papers. During this process, the most frequent topics were extracted to form the categories. The most frequently appeared topics in the literature were the approaches used, social media platforms, social media data types and countries. These categories formed the base of the mapping. To add more value to the results, we added the timing attribute. By analysing categories according to time and trends, we discover the patterns in the researches. Figure 3 shows the classification scheme that will be followed in analysing the researches.

Data extraction and systematic mapping

In this stage, the primary study set (Online Resource 1) will be screened in full for data extraction. The primary set (the 74 papers) was identified by using the first screening phase depending on the title and abstract and the second phase depending on the full-text. The data extraction process was done using Excel. The Excel sheets can be provided if needed.

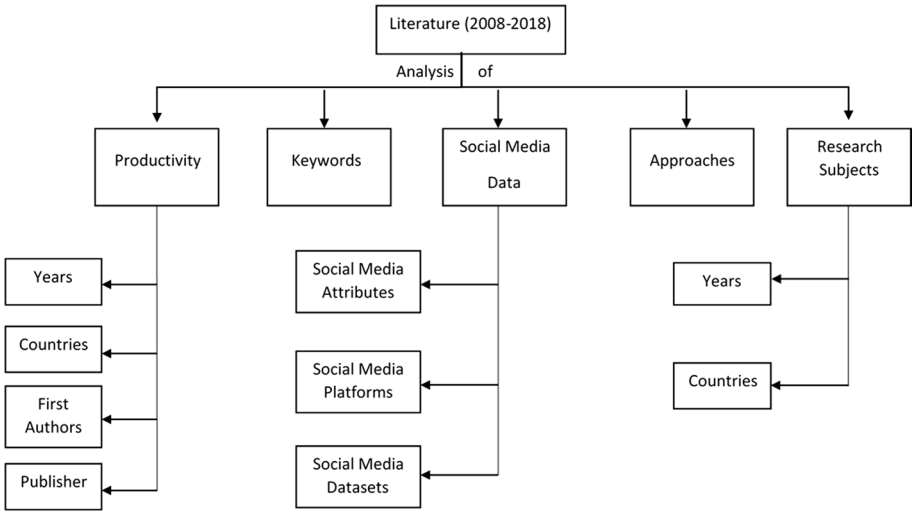


Fig. 3 The classification scheme

The data were extracted and mapped to the specified categories in the classification scheme. The results of the mapping and the analysis were visualized using Tableau which is a visualization tool widely used in the business intelligence industry.

Results

This section lays the answers to the research questions. Tableau was used for visualizing the results.

RQ1: How social media is used in transportation research based on social media analysis?

Five sub-questions were generated from RQ1 as enlisted in Table 2. By combining their answers and drawing a conclusion, RQ1 will be answered.

s1-RQ1: What is the distribution of the researches in terms of activity?

To perform an activity analysis, we have dissected the number of publications according to years, published countries, first authorship and publishers (journals and conferences).

- Activity according to years: In the years between (2008–2018), 74 papers were expurgated according to our search methodology ("RQ1: How social media is used in transportation research based on social media analysis?" Section). Our analysis points to growing attention toward the TRR-SMA field. In the period between 2008 and 2012, there was no activity in the research according to our primary set. This is believed to be attributed to the social media evolution that began in 2011–2012. (O'Regan, 2018). In 2012, Facebook users reached more than 1 billion and in 2011, the number of tweets

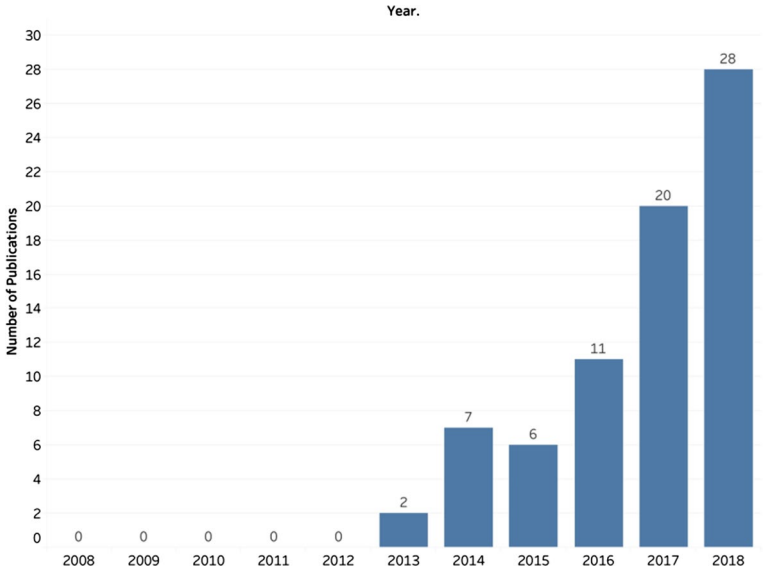
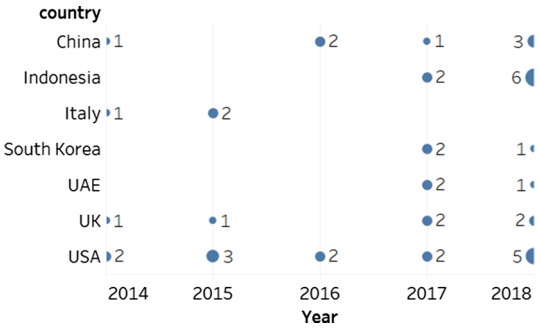


Fig. 4 The activity trend of TRR-SMA Field over the past decade

Fig. 5 The activity distribution of the top active countries



per day exceeded 140 million. In 2013, a modest activity started in the field followed by a prominent rise in the number of publications in the following years. Figure 4 illustrates the trend in the publication activity over the past decade.

- Activity according to countries: Our analysis seeks to distinguish the most interested countries in the TRR-SMA field. Figure 5 shows the top active countries in the field. The USA produced the largest share of the publications. It published around 19% of the total publications (74 papers), followed by Indonesia, China and the UK. They outputted approximately 11%, 9.5%, 8% of the publications, respectively. Cumulatively, these four countries produced around half of the publications in the field. Noticeably, the USA is dedicating high attention to the field. This high attention can be referred to the need for improvements in transportation infrastructure as stated by the Council on Foreign Relations[1] (cfr). It also stated that the USA transportation lost to South

[1] https://www.cfr.org/backgrounder/transportation-infrastructure-moving-america.

Table 7 The top publishers in the TRR-SMA field

Publisher name	Number of publications
IEEE conference on intelligent transportation systems	5
IEEE Transactions on Intelligent Transportation Systems	3
IEEE International Conference on Big Data Computing Service and Applications	2
IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining	2
Industrial Management and Data Systems	2
International Conference on Information and Communication Technology	2
Transportation Research Part C: Emerging Technologies	2

Korea, Spain and Oman transportation in 2011. USA transportation ranking retreated from fifth in the world in 2002 to twenty fourth in 2011 according to the World Economic Forums' Rankings. In addition, according to the United States Department of Transportation (TOD), in 2009, grants for transportation development were listed in the American Recovery and Reinvestment Act under the name Transportation Investment Generating Economic Recovery (TIGER). TIGER focuses on environment, energy and surface transport. In 2012, the USA president approved to progress the plan.[2] This may explain the reasons for the United States' activity, which began in 2014, and the increase in the ranking of the United States' transportation infrastructure, which was ranked sixth in the world in 2017–2018.[3]

- Activity according to first authorship: This type of activity analysis aims to discover the most active authors in the field. The activity of the authors was calculated by counting his/her publications where he/she was the first author of the publication. The top five authors in the past ten years were Gal-Tzur, Candelieri, Ali, Salas and Serna with 2 publications for each.
- Activity according to publishers: Here, the term "publishers" refers to journals and conferences that have published work in the field of TRR-SMA. Table 7 shows the top publishers and their corresponding number of publications.

s2-RQ1: What are the used keywords in the field?

Authors usually use keywords to point out the research subjects, fields, tools, approaches and techniques they use in their literature. Hence, keywords are the best identification of literature. As for other researchers in the field, they use keywords to retrieve related literature to their field from digital databases (Sharma & Mediratta, 2002). Sharma and Mediratta describe the keywords as "the "keys" to unlock the desired scientific paper abstracts/full articles from a vast collection of related publications". Due to these reasons, keywords analysis, on one hand, is important to authors as they should choose the proper keywords to identify their work and make it easy to reach. On the other hand, keywords are important to the researchers in the field to choose the suitable ones during the searching

[2] https://www.transportation.gov/50/timeline/accessible.

[3] Available on http://reports.weforum.org/pdf/gci-2017-2018-scorecard/WEF_GCI_2017_2018_Scorecard_GCI.A.02.01.pdf.

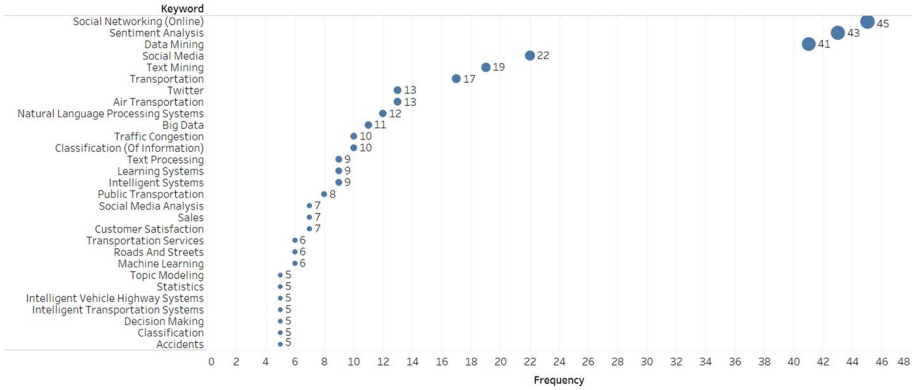


Fig. 6 The most used keywords

process to maximize the relevance of the retrieved publications. Figure 6 shows the most used keywords with their frequency.

As expected, the most used keywords are related to social media analysis and transportation fields as they are the main two topics of this paper, in addition, they were the base of the used keywords in the search process. The most used keyword "Social Networking (Online)" indicates the social media networks while "Data mining", "Text Mining", "Big Data", "Natural Language Processing Systems" and "sentiment analysis" are forming the sub-fields of social media analysis.

Other keywords are related to the approaches, data attributes, social media platforms and subjects used in the field; "Twitter" is the most used social media platform in the field (see s5-RQ1 answer); "Air Transportation", "Traffic Congestion", "Transportation Services", "Customer Satisfactions" and "Accidents" are laying under the most targeted research subjects in the field (refer to RQ2 answer).

s3-RQ1: What are the social media data/attributes used by the researchers?

One of the pillars for structuring the TRR-SMA area is the data used in the studies. Users' posts on social media sites will provide data other than the text, picture, or video they posted. In addition to time, users' posts will be tagged with the location if they activate the geotagging feature. Online Resource 1 shows the used social media data by the researchers. Text, location and time data were the frequently- used social media attributes.

Other data such as followers, number of tweets, friends and retweets were used to gauge the users' influence. The number of tweets was also used as a gauge of traffic congestion. In terms of ratings, they were used as a gauge of public opinion.

s4-RQ1: What are the rules of text data/text mining in the TRR-SMA field?

Text data was utilized by all papers in our primary set. Further investigation on the role of text data and the used attributes of text data was made and presented in Table 8. Location and time attributes are important in the transportation domain especially in identifying road conditions and issues; they are fundamental in defining the exact location and time of incidents and traffic. Therefore, social media is considered as a real-time source of

Table 8 Roles of Text Data in TRR-SMA Field

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification				Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Abalı et al., 2018)	√															
(Alamsyah et al. 2018)										√						
(Alamsyahl & Rachmadiansyah, 2018)																
(Ali et al., 2018)		√														
(Buch et al., 2018)		√														
(Y. Chen et al., 2018)										√						
(Fiarni et al., 2018)		√		√						√						
(Gal-Tzur et al., 2018)									√						√	
(Gupta et al., 2018)		√														
(Haghighi et al., 2018)		√								√	√					
(Hosseini et al., 2018)										√						

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification			Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Kaur & Balakrishnan, 2018)			√	√		√	√		√							
(Kovács-Győri et al. 2018)		√														
(Kulkarni et al., 2018)		√								√						
(K. Lee & Yu, 2018)		√														
(Musaev et al., 2018)																
(Rane & Kumar, 2018)																
(Rybarczyk et al., 2018)		√										√				
(Salas et al., 2018)	√															
(Samonte et al. 2018)																

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification			Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Key-words	Origin–Destination	Mentions
(Saragih and Girsang 2018)		√						√								
(Sdoukopoulos et al. 2018)		√								√						
(Serna & Gasparovic, 2018)												√				
(Sternberg et al., 2018)										√						
(C. Wang et al. 2018)																
(Wayasti et al., 2018)																
(Windasari et al., 2017)				√				√		√						
(Z. Zhang, Chen, et al., 2018; Zhang, Zhang, et al., 2018)																

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification				Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Ali et al., 2017)		√								√						
(AlSheikh et al., 2017)																
(Anastasia & Budi 2016)										√						
(Baj-Rogowska, 2017)		√														
(Casas & Delmelle, 2017)	√															
(Dutta Das et al., 2017)																
(Kuflik et al., 2017)										√						
(Lu et al. 2017)										√						
(Luckner et al., 2017)	√															
(Pournarakis et al. 2017)										√	√					√

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification			Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Salas et al., 2017)	√	√														
(Saldana-Perez et al., 2017)																
(Septiana et al., 2016)	√															
(Serna et al., 2017)												√				
(Sinha et al.. 2017)	√									√					√	
(Suma et al., 2017)										√						
(Thelwall, 2017)		√		√	√		√	√	√							
(D. Wang et al., 2017)	√															
(L. Zhang et al., 2017)																
(Kim et al. 2017)																

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification				Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(S. Chen et al., 2016)		√								√						
(Gao et al., 2016)										√						
(Giancristofaro & Panangadan, 2016)											√					
(Hoang et al., 2016)		√								√	√					√
(Itoh et al., 2016)	√									√						
(Lacic et al., 2016)																
(Liyang et al., 2016)										√	√					√
(Tse et al., 2016)										√						
(Ulloa et al., 2016)		√								√						

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification			Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Yang & Anwar, 2016)		√						√		√						
(B. Zhang, Kotkov, et al., 2016; Zhang, Sun, et al., 2016)		√														
(Candelieri & Archetti, 2015)										√		√			√	
(D'Andrea et al., 2015)										√						
(Fu et al., 2015)										√						
(Georgiou et al., 2015)	√									√						
(Rahman et al., 2015)																
(X. Zhang et al., 2015)																

Table 8 (continued)

Text Role	Location Detection	Sentiment Analysis								Topic Extraction/ Classification				Weather Detection	Routs Detection	Account/ Company Detection
Study	Location names	Sentimental words	Capitalization	Negation	Boosters	Intensifiers	Letter Repetition	Emotions	Punctuation	Keywords/ Terms	Hashtags	Transport Mode	Features/ Aspects	Keywords	Origin–Destination	Mentions
(Adeborna & Siau, 2014)		√														
(Candelieri & Archetti, 2014)																
(Cao et al. 2014)		√		√		√										
(Carpenter et al., 2014)		√														
(Gal-Tzur et al., 2014)										√						
(Kumar et al., 2014)	√									√						
(Liau & Tan, 2014)		√		√	√		√		√							
(Daly et al., 2013)	√													√		
(Mostafa, 2013)		√									√					

information as people post incidents that have happened. Many papers did not just employ the geotagged location, as they also used text data to detect locations. This attributes to the low amount of geotagged data on social media (Sloan and Morgan 2015).

S5-RQ1: What are the social media platforms used by the literature?

Social media platforms become a major part of humans' life, and with their importance, many platforms have been unveiled. Choosing the proper platforms for research is essential. Online Resource 1 shows the social media platforms used by researchers in the TRR-SMA field and Fig. 7 shows the usage trend of the platforms over the past decade. Surprisingly, the platforms' usage trends are divisive; on the one hand, Twitter is the most popular platform in the field, despite not being the most popular worldwide. On the other hand, while Facebook is the most popular social media platform worldwide, its use in the field is limited. In a ten-year period, it was cited in just 9 out of 74 reports. Twitter is the data source for approximately 72% of the papers. It is a microblog; it limits the number of characters per tweet and tags the tweets with the time and location (if users allowed). Facebook offers users the same features with one main difference—there is no limit on the number of characters per post. Due to this, it is thought that the processing of tweets is easier than Facebook posts. As the number of tweets characters are limited, people will directly point to their subject without further explanation or description.

However, the main reason behind the limited usage of Facebook data is retrievability. Facebook announces limitations on its APIs[4] (application programming interface). Facebook APIs are used to crawl Facebook data, and restricting API access means limiting Facebook data access. In addition, Twitter has announced that precise location tagging would be removed from their platform.[5] According to the company, the precise location tagging will be available for images taken with Twitter's camera. This creates a challenge for fields that depend on precise location, such as transportation, and necessitates the search for alternate ways to detect locations.

S6-RQ1: What are the datasets used by researchers?

The contents of social media platforms are either open or closed; some platforms allow content retrieval through APIs, while others do not. The datasets used by researches were collected as follows:

- Twitter datasets: Twitter datasets were collected using the APIs. Twitter APIs allow the developer to access and retrieve Twitter contents including users' data and timeline, retweets, hashtags data and others. The number of allowed queries and retrieved tweets from users' timeline differs according to the API type.[6] Two main libraries are used to extract tweets: Twitter4J for Java and Tweepy for Python.

[4] https://techcrunch.com/2018/07/02/facebook-rolls-out-more-api-restrictions-and-shutdowns/.

[5] https://twitter.com/TwitterSupport/status/1141039841993355264?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1141039841993355264&ref_url=https%3A%2F%2Fwww.theverge.com%2F2019%2F6%2F19%2F18691174%2Ftwitter-location-tagging-geotagging-discontinued-removal.

[6] Twitter developer APIs: https://developer.twitter.com/en/docs/basics/getting-started.

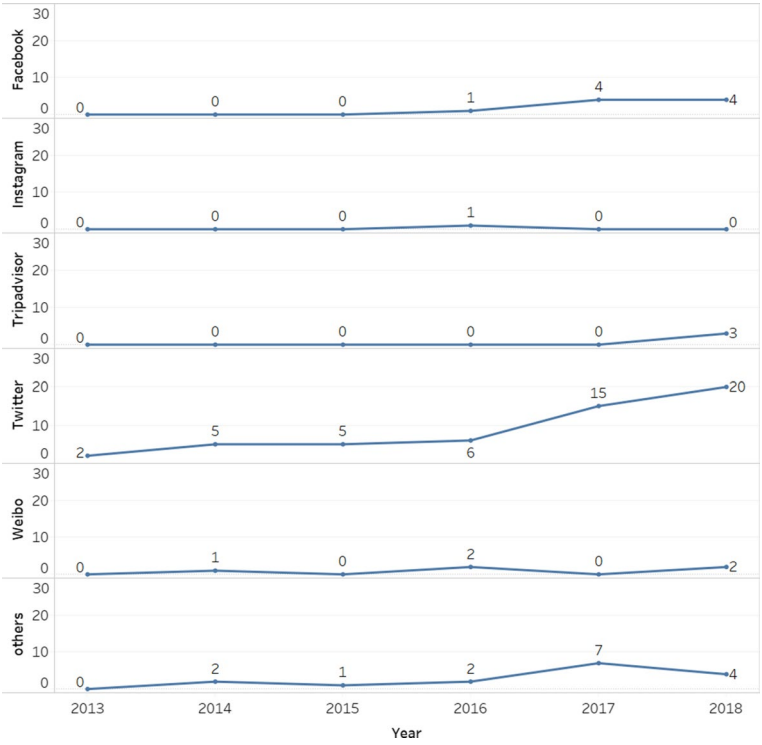


Fig. 7 The usage trends of the social media platforms

- Facebook datasets: Mainly Graph API used by researchers to collect Facebook data. The problem with the Graph API[7] and other Facebook APIs that Facebook limits the access to the users' data.[8] ProSuite tool to collect Facebook data was used by (Baj-Rogowska, 2017) while (Ali et al., 2017) collected the data manually.
- Weibo datasets: There are two approaches to retrieve Weibo contents which are the requesting APIs and crawling. APIs usually are paid and limit the number of queries (Y. Chen et al., 2018). Crawling is performed using HTTP request (Y. Chen et al., 2018) or crawlers such as Crawlzilla and Selenium (S. Chen et al., 2016)
- TripAdvisor datasets: The TripAdvisor content APIs are available, but they are only for use on travel websites. Since TripAdvisor's APIs are not available for academic research or data analytics,[9] scrappers and crawlers are used to obtain the data.
- Others: other datasets from review platforms such as Google reviews (K. Lee & Yu, 2018) and Yelp (Gao et al., 2016) are used. Google reviews can be retrieved using Google APIs.[10] Google APIs have a retrieval limit of 5 reviews per location. As for Yelp, it offers an open dataset for academic purposes[11] besides its APIs for business.

[7] Facebook Graph API: https://developers.facebook.com/docs/graph-api.
[8] https://developers.facebook.com/policy/.
[9] TripAdvisor API access policy: https://developer-tripadvisor.com/content-api/request-api-access/.
[10] Google API: https://developers.google.com/android-publisher/api-ref/reviews.
[11] Yelp dataset: https://www.yelp.com/dataset.

s7-RQ1: What are the approaches used to analyse social media data in transportation researches based on social media analysis?

The second step, after defining the data that will be used, is to understand how to use it to obtain the desired useful information. Machine learning-based (ML) approaches, natural language processing-based (NLP) approaches, and statistical-based (SL) approaches are the three groups in which these approaches were classified. These categories, however, can overlap since several methods fall into more than one of the three categories.

Machine learning algorithms are commonly used for classification, regression, and grouping related instances; in the literature, they were mostly used to identify the commuters' perspective toward the transportation network or to identify the related posts to transportation or events. Machine learning approaches can be divided into two categories: supervised and unsupervised.

Supervised machine learning methods need pre-labelled data as training samples to perform classification. Support Vector Machine (SVM) and Naïve Bayes (NB) are ones of the most popular supervised machine learning approaches. They are well-known for their performance in text classification field; hence they have been commonly used in the literature. The SVM algorithm maximises the distance between training data groups and draws a hyperplane between them. Then, it decides which side of the hyperplane the new instances belong to by using the features of the new instances and the information obtained from the training data. SVM was employed by studies for sentiment analysis to identify the public opinion regarding the transportation network or their complaints (Candelieri et al., 2015; Pournarakis et al., 2017; Sinha et al., 2017; Windasari et al., 2017; Yang et al., 2016) and for distinguishing the transport-related posts from the irrelated ones (Y. Chen et al., 2018; Gal-Tzur et al., 2014; Salas et al., 2017; Salas et al., 2018). NB utilises Bayes theorem from statistics. Using the training data and the features of the new instance, NB calculates its likelihood to belong to a class. NB was employed to classify the related posts to transportation or events by (Abalı et al., 2018), to analyse commuters' sentiment toward transportation by (Alamsyah et al., 2018; Dutta Das et al., 2017; Fiarni et al., 2018; Kumar et al., 2014; Liyang et al., 2016; Sternberg et al., 2018) and to predict vehicle recall by (X. Zhang et al., 2015). Multiple researches compared the two, and some of these researches compared them to other approaches including decision trees (DT) and for the same previous aims; (Alamsyahl et al., 2018; Anastasia & Budi, 2016; Giancristofaro et al., 2016; Gupta et al., 2018; Rane & Kumar, 2018; Z. Zhang, Zhang, et al., 2018; Zhang, Chen, et al., 2018) used multiple classifiers to compare their results in analysing commuters' sentiment toward transportation related topics, while (D'Andrea et al., 2015; Gal-Tzur et al., 2018; Hoang et al., 2016; Kuflik et al., 2017; Tse et al., 2016) used different machine learning techniques to identify the posts related to a transportation topic. Other classification approaches such as Maximum Entropy (ME) (Dutta Das et al., 2017; Samonte et al., 2018) and Logistic Regression (LR) (Rane & Kumar, 2018; Zhang, Chen, et al., 2018; Zhang, Zhang, et al., 2018) were also used by some researchers. ME selects the appropriate distribution to represent the data based on the measurement of entropy, whereas LR is used to represent the data and describe the relationship between variables.

In regard to the unsupervised machine learning methods, they do not need pre-labelled data; instead, they rely on data features to find the similarity between instances. K-Nearest Neighbour (KNN) is unsupervised machine learning algorithm which is used for clustering (D'Andrea et al., 2015; Kumar et al., 2014; Rane & Kumar, 2018; Saldana-Perez et al., 2017; X. Zhang et al., 2015; Z. Zhang, Zhang, et al., 2018; Zhang,

Chen, et al., 2018). Another popular clustering algorithm, which is used within the sentiment analysis process to discover the top topics, is k-means and its version spherical K-means (Liau & Tan, 2014) where both of them produced similar topics.

Another approach of ML is Deep learning (DL). DL is known for its high accuracy and its ability to learn from large amount of data (Kim et al., 2017). DL includes many architectures, one of the basic architectures is Multilayer perceptron (MLP) which was used by two studies (Ali et al., 2018) for sentiment analysis purpose and by (Chen et al., 2018) for identifying the traffic-related information. Chen et al. (2018) used a combination of other DL architectures, namely: convolutional neural networks (CNNs) and long short-term memory (LSTM) and compared them with other machine learning methods.

As text data is the most commonly used data in the research, natural language processing-based approaches have a dominant rule in the information extraction and data analysis processes. Before performing any data analysis/classification or knowledge extraction task, the bag of words (BOW) approach is typically used to represent text data. BOW converts text data into a numerical form that ML algorithms and others beside computers can understand. (Chen et al., 2018; Gal-Tzur et al., 2014; Giancristofaro et al., 2016; Liau & Tan, 2014; Musaev et al., 2018; Pournarakis et al., 2017; Rane & Kumar, 2018). N-gram is a model under computational linguistics and refers to sequence extraction from text or speech, so generating n-grams is included in the pre-processing stage of texts in the studies (Ali et al., 2018; Daly et al., 2013; Windasari et al., 2017). Another pre-processing stage of text is parsing, parsing indicates the syntax analysis of text data and usually is used to extract the grammatical rules of the language (Luckner et al., 2017; Zhang, Kotkov, et al., 2016; Zhang, Sun, et al., 2016). The dominant natural language processing approach in the studies is lexicons. This likely to be due to its simplicity. There are two types of lexicons: the sentimental lexicon (SL) and the dictionary (Dic). The dictionary contains the words related to domain (domain lexicon) or language (general lexicon) and a sentimental lexicon is a dictionary which associates each word with its sentiment polarity. The generation methods of lexicons may include interference of machine learning or/and statistical-based approaches as it is shown in Table 9. Table 9 illustrates the lexicons used by researchers and the generated ones. Commonly, Bing Liu lexicon (Hu et al., 2004) is the most used general lexicon.

Another approach which can be used in the pre-processing stage of text data is Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a statistical-based information retrieval metric and it is one of the most common and traditional term weighting approaches in which it assigns a weight for each token in the text relying on its occurrences. TF-IDF was used by the researchers mostly to weight the words in the topic modelling process (Candelieri et al., 2014; Fu et al., 2015; Itoh et al., 2016; Sinha et al., 2017; C. Wang et al., 2018; Windasari et al., 2017; Yang et al., 2016; Z. Zhang, Zhang, et al., 2018; Zhang, Chen, et al., 2018). Topic modelling approaches usually use statistical models to represent the data and draw inferences. Latent Dirichlet Allocation (LDA) is the most popular topic modelling approach. It uses Dirichlet distribution to represent the text data to find the most important topics. It was used by 13.5% of the studies (Alamsyah et al., 2018; Buch et al., 2018; Kovács-Győri et al., 2018; Kulkarni et al., 2018; Lee et al., 2018; Pournarakis et al., 2017; D. Wang et al., 2017; Wayasti et al., 2018). Researchers used LDA to identify sub-topics that could signify aspects of the transportation network that users discussed or issues that users encountered. LDA was used before classification algorithms to find categories/topics since the result of LDA could present categories.

RQ2: What are the aims of transportation researches based on social media analysis?

To identify these subjects, the sub research questions of RQ2 are answered as follows:

s1-RQ2: Which subjects were targeted by researchers in the TRR-SMA field?

Transportation is a broad subject; this leads the researchers to target the transportation subject in general or one or more of its sub-subjects. The works in our primary set are classified according to their target as illustrated in Fig. 8. Online Resource 1 shows the literature and their targets. It groups the literature according to the targeted countries to demonstrate the trend and the targeted subjects.

s2-RQ2: What are the trended subjects in the world and by countries?

To demonstrate the trend of the research subjects in each country, the table in Online Resource 1 groups the literature according to the targeted countries. The target country is the source location of the data. The source location is defined when retrieving the data using the APIs. The Worldwide location was assumed in the case of the undefined location of the data, except for Air Transport subject, it was assumed worldwide if the work targeted multiple airlines from different countries. Otherwise, the source location will be the airline's home country. Figure 9 illustrates the distribution of the subjects over the past decade in the world.

From Online Resource 1 and Fig. 10, we can extract the dominant subjects and their trends in each country; in the USA, the dominant subject was "general". The subject "general" was trending from 2016 till 2018 while "Issues" was trending in 2015. "Road Transport" has a little interest; in the UK, the "Issues" and "General" subjects got equal attention, the "General" subject was trending in 2018 and "Issues" was trending in 2017–2018. We can say the UK has no interest in Air Transport as no publications targeted it; in China, the concentration was forwarded to the "Issues" subject, it was targeted by researchers in 2014, 2016 and 2018; in Indonesia, the dominant target was "OT", it was the target of 6 papers out of 7, this can indicate the weakness of the transportation network that leads the commuters to depend on OT services; in other countries, the trend is hard to be tolled as they were targeted by a small number of researches.

However, other information can be extracted such as the region of services. Air Transport is a global service. This is logical as airlines responsible for moving people through countries. Another interesting fact is regarding OT services—OT-Uber was the target when worldwide was the location and this can indicate that Uber[12] is a global company. Grab and Gojek were the target when the researches targeted Indonesia. This indicates that Grab[13] and Gojek[14] are local or at most regional companies. This fact can be proved by looking at the companies' websites.

[12] Uber locations: https://www.uber.com/global/en/cities/.
[13] Grab locations: https://www.grab.com/my/locations.
[14] Gojek: https://www.gojek.com/about.

Table 9 Used and generated lexicons analysis

Literature	Used lexicon type	Used lexicon/ tool name	Generated lexicon type	Generation method
(Adeborna & Siau, 2014)	SL-General	Bing Liu	SL-Domain	Modifications to the lexicon by adding domain-dependent words from WordNet and tweets using Correlated Topics Models (CTM) with Variational Expectation–Maximization (VEM) based on Airline Quality Rating (AQR) criteria resulted in 4 lexicons concerning 4 topics related to airlines
(Ali et al., 2017)	SL-General	SentiWordNet	SL-Domain	SentiWordNet for scoring with Fuzzy-Ontology and Semantic web rule language (SWRL) for rule-based decision-making
(Baj-Rogowska, 2017)	SL-General	CAT	–	–
(Buch et al., 2018)	SL-General	SentiStrength	–	–
(Cao et al., 2014)	SL-General	HowNet	SL-Domain	Generating lexicon by expanding seed words for Chinese
(Carpenter et al., 2014)	SL-General Dic-General	MPQA WordNet	SL-Domain	Using WordNet to expand the seed words through their synonyms
(Fiarni et al., 2018)	–	–	SL-Domain	Generated using rule-based approach with NB considering negation
(Gal-Tzur et al., 2014)	–	–	Dic-Domain	Constructed using 35 documents from research articles, websites and forums by term frequency and specialists
(Gupta et al., 2018)	SL-General	Bing Liu, MPQA, AFINN, SentiWordNet	–	–
(Haghighi et al., 2018)	SL-General	Rsentiment Package	–	–
(Hosseini et al., 2018)	–	–	Dic-Domain	Using 2 Thesaurus, Transportation Research Thesaurus (TRT) and the Australian Transport Index Thesaurus
(Kaur & Balakrishnan, 2018)	SL-General		–	Used general lexicon with letter repetition, intensification, capitalization, negation and exclamation mark to calculate the sentiment of the words

Table 9 (continued)

Literature	Used lexicon type	Used lexicon/ tool name	Generated lexicon type	Generation method
(Kulkarni et al., 2018)	SL-General	VADER	–	–
(K. Lee & Yu, 2018)	SL-General	AFINN	–	–
(Liau & Tan, 2014)	SL-General	Bing Liu	SL-General	The Malay lexicon was created manually
(Mostafa, 2013)	SL-General	Bing Liu	–	–
(Rybarczyk et al., 2018)	SL-General	ANEW	–	–
(Salas et al., 2017)	SL-General	SentiStrength, TensiStrength	–	–
(Saragih and Girsang, 2017)	SL-General	Bing Liu, library of sentiment for Indonesian words	SL-General	The Indonesian lexicon was generated by translating Bing Liu lexicon
(Sdoukopoulos et al., 2018)	SL-General	NodeXL	–	–
(Serna et al., 2017)	Dic-General	WordNet	SL-Domain	WordNet used to expand word related to transportation and then used the total rank of the text to define the positive and negative words
(Thelwall, 2017)	SL-General	TensiStrength	–	–
(Yang & Anwar, 2016)	SL-General	SentiWordNet	–	–
(L. Zhang et al., 2017)	SL-General Dic-General	-VADER -WordNet	SL-Domain	Define seed set manually then extend it through wordnet iteratively. They used VADER to assign a level of sentiment
(Daly et al., 2013)	–	OpenStreet Maps	Dic-Location	–
(Kuflik et al., 2017)	–	–	Dic-Transport	35 transport documents were used: stakeholders' Web sites (e.g. of taxi services, transport magazines, etc.); research articles and white papers, transport Web forums, blogs and SM accounts
(Saldana-Perez et al., 2017)	–	–	Dic-Traffic	Using specified classes and TF to build traffic-related dictionary from tweets
(Tse et al., 2016)	–	–	Dic-Pollution	Most frequent keywords in the related posts

Table 9 (continued)

Literature	Used lexicon type	Used lexicon/ tool name	Generated lexicon type	Generation method
(D. Wang et al., 2017)	Dic-General Dic-Location	Google twitter frequent words—British Telecommunications (BT) Dictionary	–	–

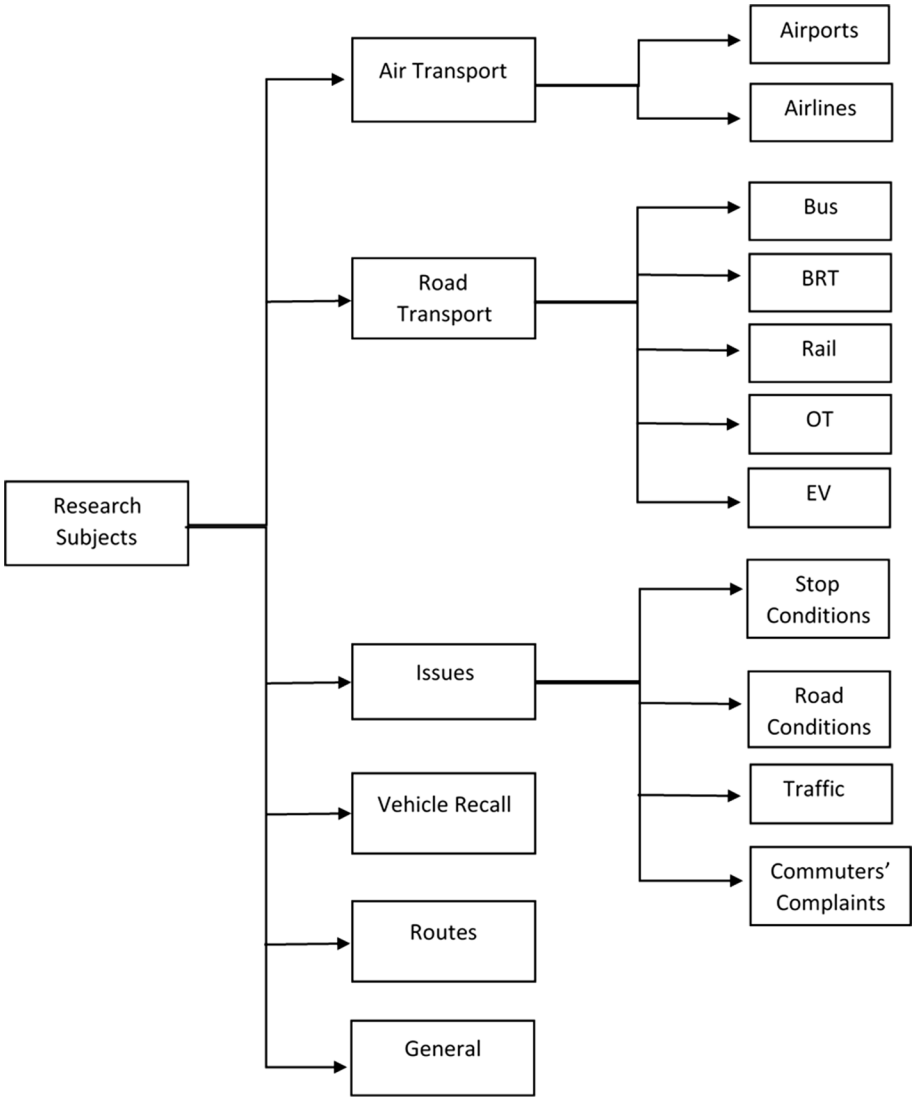


Fig. 8 Targets classification. OT = Online Transport, EV = Electric Vehicle, BRT = Bus Rapid Transit

s3-RQ2: What are the social media attributes used for achieving the targets?

To draw the answer to this question, the subjects were associated with the attributes used (see Online Resource 1). Noticeably, text, location and time were mostly used together for exploring the traffic situation and causes, roads and, in general, network conditions. Text data and ratings were utilized for measuring the commuters' attitude

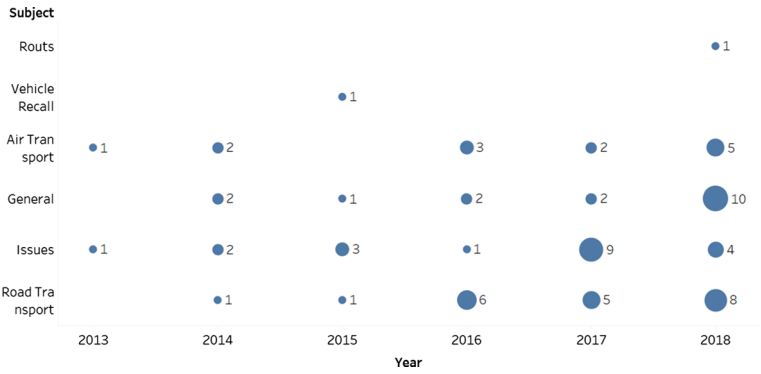


Fig. 9 The trended subjects around the world in the past decade

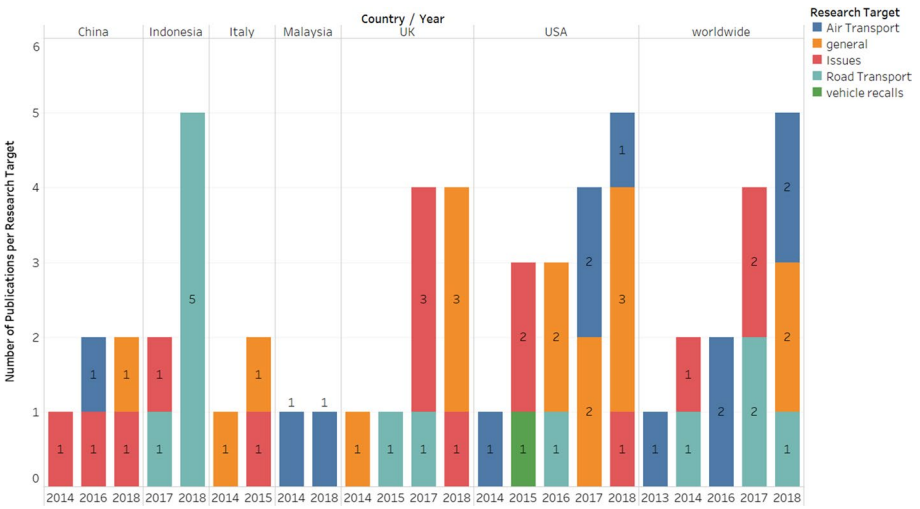


Fig. 10 The trended subjects in the top locations

towards the transportation network. However, text data is the base. It can be employed to fulfil any target.

RQ3: What are the challenges, principal findings and possible future works in the field?

Concluding the answers of the previous RQs, we can recognize challenges and future works. Challenges and future work are drawn from the findings as follows:

- Activity: Through activity analysis (see s1-RQ1 answer), the evolution trend of the TRR-SMA field was shown. The activity in the field is increasing through time. The country with the dominant role in the production process is the USA with 14 publications and the transport-related publishers are the biggest source of the publications. A

related challenge to activity analysis is to demonstrate the reasons behind the inactivity or small activity of countries such as Japan, Turkey, Singapore, Canada and others. This could be referred to the economic situation, the producibility of researches, the availability of the data, the interest in the transportation infrastructure and the priority of transportation planning and enhancement and many others. An in-depth review of the grey and white literature (organisations reports and technical documents) in addition to official news about the development of the transportation network may identify these reasons.

- Social Media Data: The attached Online Resource 1 shows each study of the primary set with its corresponding subject, the data used by the researchers to achieve their aims and the platform used. Table 8 shows the purpose of using the text data, and the text attributes used for each purpose. By combining Online Resource 1 and Table 8, the subject of the research with the corresponding data used and the purpose of using text data can be extracted for each study. Furthermore, it can be concluded that text was used by all papers for different goals owing to the fact that text is the main content of the shared posts on social media platforms such as Twitter, Facebook and Weibo. Further abilities of text data need to be explored like providing precise data about weather situations and location. As less than 1% of the social media data is geotagged, extraction of location becomes one of the most important objectives of text data; however, there are several issues surrounding the extracted location from text, such as whether the extracted location is the event/incident location. Is the place that was extracted a commuter location? Is the extracted name a location name? Is the place that was extracted fake? To find answers to these questions, further research is needed. As for weather, it can seriously affect the transportation network as it can cause closures of roads and others. One of the research in our primary set extracted weather from text using keywords (Daly et al., 2013) and another used other resources to get weather data (Rybarczyk et al., 2018). According to the impact of weather on transportation, further consideration by researchers is needed besides exploring the efficiency of text data in providing precise weather information. In other words, how likely would people share the weather information through social media and to which degree is the extracted data true and precise. Another goal of text is to extract transportation modes (e.g. metro, bus, taxi); however, more research is needed because modes may have different names in different regions or even countries.
- Other essential data in transportation studies is time. Time is easy to retrieve as it is tagged to posts as people share but to which extent it had been explored. In general, time is used by researchers in identifying incidents time and traffic time. However, other usages of time need to be explored such as delay detection and validation of extracted location from text, as several locations from different posts can be extracted and compared according to time to see if it is possible to move between these locations during the recorded time.
- Social Media Platforms: Multiple applications were used as sources of the data. Many researches used multiple applications as a source (refer to Online Resource 1). Facebook data was not used frequently. This perhaps is related to the limitations on Facebook APIs and its allowance of long posts. One of the concerns that may face the researchers in the future is Twitter ending its exact location geotagging service and limiting it to images taken by its camera. This will open the need for further research using images and text to detect locations or even exploring the efficiency of other social media applications such as Instagram. Instagram can reflect the real situation of

transportation network since its main content is images, so Instagram and, in general, images efficiency in delivering transport-related data needs further investigation.

- Social Media datasets: In general, self-extracted datasets were used in the researches; each group of researchers collected their own dataset from the platforms using either the APIs or the crawlers/scrappers. This makes it difficult to compare the results of publications in the same subject. Two key factors affect the transportation domain are real-time data and location; hence, by using different query attributes such as location and time, different datasets can be retrieved. However, a standard open dataset is required so the performance of the research can be compared.
- Approaches: Lexicons, SVM and NB are the most employed approaches in the TRR-SMA field. The usage of lexicons is the highest. The high usage can be related to the lexicon's simplicity and effectiveness in topic and sentiment classification. The problem with lexicons is domain dependency, coverage and outdating. Domain dictionaries that are used for topic classification have a problem in detecting the domain words, usually LDA, TF or TF-IDF will be used in the detection process, however, these methods will also result in a set of unrelated words to the domain. As a result, manual filtration of the domain-related words is used by researchers which needs time and effort. This rises to surface the need for automatic filtration of the words. Other problems are with sentimental lexicons. The coverage and outdating of words are issues of many lexicons. Take for example the most used general lexicon by literature—Bing Liu (Hu & Liu, 2004). It composes of approximately 7000 words, while the Oxford dictionary contains around 170 K words.[15] This big difference in the number of words creates the coverage issue. Another issue is the updates of lexicons and the usage of words through time where the new generation may stop using some words and start using other words to indicate other meanings (Schulz et al., 2010). To overcome the previous mentioned issues, frequent updates of lexicons are needed, yet how frequently are the general lexicons being updated? Moreover, the sentiments of words may change depending on the domain or context, hence recognizing the changeable sentiment is a dilemma for general lexicons. To the extent of the authors knowledge, there is no transport-related lexicon. The creation of this lexicon is believed to improve the performance of the systems that uses transportation related data.
- Subjects: Focuses or targets of researches were identified and grouped by countries. The aim of grouping was to indicate the trends and targeted subjects by country. One of the future directions can investigate why some publisher countries perform research on other countries. This can be due to huge incidents that occur in the targeted country, the requirement of funding agencies, shortages in the research field in the targeted country, availability of data, authors origin countries and others. Other possible directions are presented as follows:

Transportation modes: Air transport and road transport modes were the most explored in the researches, this likely due to the availability of data and the availability of the transportation modes on most of the countries. On one hand, most air transportation research has focused on customer opinion mining to investigate service quality by gathering posts and feedback on the airlines' or airports' social media pages. On the other hand, research on road transportation focuses on finding out what commuters think about rail, OT, and

[15] https://en.wikipedia.org/wiki/List_of_dictionaries_by_number_of_words.

buses, as well as issues of transportation infrastructure. However, other research directions can be explored regarding the two, air transport and road transport, such as creating automated alerting system in case of delays or accidents or automated reply system to commuters' inquiries and complaints. Other modes of transport are water transport modes such as ferries. Even though water transport is a significant mode of transportation in many countries, no research has focused on it. Water transport was part of the general subject in a few studies (Gao et al., 2016; Rybarczyk et al., 2018).

Routes: In routes, the origin destination locations were extracted from text. Usually, the combination of "from" and "to" is used for this purpose. In certain cases, users will simply mention the destination; in these cases, a method to find the origin, as well as a method to verify that the extracted location is actually a destination are required.

Issues: In the event that an issue with the transportation network arises such as road hazards, accidents, road closures and others, social media may serve as an early warning device. However, social media posts are human-created knowledge that is not always accurate and can be influenced by people's moods and psychology. Validation methods for the extracted issues are needed, particularly in the event of an emergency or sudden problem. One of these methods is comparing the extracted information from social media with official news and other potential resources taking into account the time and location of the extracted information. However, further research into text analysis and summarisation techniques is needed for the comparison purposes. Furthermore, further research is required on automated issues discovery from text rather than relying on human effort or pre-defined lexicons.

Recommendation systems: The studies did not cover transportation recommendation systems. Personalisation is an attribute that social media may provide to recommendation systems. It can be used in OT services to suggest personalised rides, especially in ride-sharing services where multiple passengers can ride together. The ride can be recommending based on the profile of the passengers to assure more friendly and personalised rides. In other cases, social media may be used to suggest routes, especially in the event of unexpected closures or extreme traffic congestion. Furthermore, by analysing users' activity patterns, personalised trips/routes can be suggested. Moreover, social media can employ the public trend in a particular location (e.g. city) to recommend transportation modes which will not be recommended by transportation application such as scooters and cabriolets.

COVID-19: During the COVID-19 pandemic, several countries imposed movement restrictions, resulting in a significant drop in traffic and, in some cases, empty ways. Furthermore, at times, all modes of transportation were shut down, and even when they were running, commuters' complaints were different than they were prior to COVID-19. Commuters' main concerns in the COVID-19 period would be travel restriction laws, COVID-19 possible transmission methods in transportation modes, and the required precautions to obey during the rides, in addition to trip cancellation and compensation. This opened the door to new research directions, such as using social media data as a warning in the event of a transportation emergency, for example: fainting persons in the transportation and discovering new COVID-19 cases in the transportation or using it as a source of information about people who do not follow COVID-19 precautionary rules during the rides such as social distancing and mask wearing, among others.

Conclusion

In this work, we structured the TRR-SMA field by performing a systematic mapping review. We identified the foundations of the field by prior reading and constructed the classification scheme based on them. In addition, the query terms were defined. These terms were used afterward to retrieve the researches from 4 DLs: IEEEXplore, ACM, Web of Science and Scopus. The search results were refined using the ECs and ICs terms. In the end, 74 papers were included in the primary set.

The foundations of the classification scheme were activity, keywords, social media data, social media platforms and targets. Through the analysis, the trends were drawn and discussed a.

Activity analysis was done in terms of country, year, publisher and first author. The TRR-SMA field is getting increasing attention. Throughout the years, the most productive country was the USA and most productive publishers were the transportation-related publishers. Moreover, publications were analysed in terms of the data, platforms and approaches used. Text data was the most utilized data by the papers. Hence, further analysis of text data was performed and presented in terms of the aims and the corresponding used text attributes. In addition, an analysis of used lexicons was presented as lexicons are the most used approach. In the end, an analysis in terms of research subjects was presented. In the analysis of the subjects, papers were grouped by countries to show the trends and the covered subjects in each country.

By accumulating and analysing the results, possible challenges and future works were drawn and discussed. These challenges and future works can guide new researchers and create new research opportunities. The most crucial future works is creating a transport-specific lexicon, creating personalised transport related recommendation systems using social media data, conducting researches regarding water transport and exploring the effect of COVID-19 on the TRR-SMA field.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-021-04046-2.

Acknowledgements This work was supported by the University of Malaya under Grant RK004-2017 and Sunway University under Grant CR-UM-SST-DCIS-2018-01

References

Abalı, G., Karaarslan, E., Hürriyetoğlu, A., & Dalkılıç, F. Detecting citizen problems and their locations using twitter data. In 2018 6th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), 25–26 April 2018 2018 (pp. 30–33). doi:https://doi.org/10.1109/SGCF.2018.8408936.

Adeborna, E., & Siau, K. An approach to sentiment analysis - The case of airline quality rating. In 18th Pacific Asia Conference on Information Systems, PACIS 2014, 2014.

Alamsyah, A., Rizkika, W., Nugroho, D. D. A., Renaldi, F., & Saadah, S. Dynamic large scale data on Twitter using sentiment analysis and topic modeling case study: Uber. In 6th International Conference on Information and Communication Technology, ICoICT 2018, 2018 (pp. 254–258). doi:https://doi.org/10.1109/ICoICT.2018.8528776.

Alamsyahl, A., & Rachmadiansyah, I. (2018). Mapping online transportation service quality and multiclass classification problem solving priorities. In International Conference on Data and Information Science (Vol. 971, Journal of Physics Conference Series). Bristol: Iop Publishing Ltd.

Ali, F., Ei-Sappagh, S., Khan, P., & Kwak, K. S. Feature-based Transportation Sentiment Analysis Using Fuzzy Ontology and SentiWordNet. In 9th International Conference on Information and

Communication Technology Convergence, ICTC 2018, 2018 (pp. 1350–1355). doi:https://doi.org/10.1109/ICTC.2018.8539607.

Ali, F., Kwak, D., Khan, P., Islam, S. M. R., Kim, K. H., & Kwak, K. S. (2017). Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. [Article]. *Transportation Research Part c: Emerging Technologies, 77*, 33–48. https://doi.org/10.1016/j.trc.2017.01.014

AlSheikh, S. S., Shaalan, K., & Meziane, F. Consumers' trust and popularity of negative posts in social media: A case study on the integration between B2C and C2C business models. In 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC), 16–18 Oct. 2017 2017 (pp. 1–6). doi:https://doi.org/10.1109/BESC.2017.8256364.

Anastasia, S., & Budi, I. Twitter sentiment analysis of online transportation service providers. In 8th International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016, 2017 (pp. 359–365). doi:https://doi.org/10.1109/ICACSIS.2016.7872807.

Baj-Rogowska, A. Sentiment analysis of Facebook posts: The Uber case. In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 5–7 Dec. 2017 2017 (pp. 391–395). doi:https://doi.org/10.1109/INTELCIS.2017.8260068.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

Buch, R., Beheshti-Kashi, S., Nielsen, T. A. S., & Kinra, A. (2018). Big Data Analytics: A Case Study of Public Opinion Towards the Adoption of Driverless Cars. In Dynamics in Logistics (pp. 347–351, Lecture Notes in Logistics). Cham: Springer International Publishing Ag.

Caetano, J. A., Lima, H. S., Santos, M. F., & Marques-Neto, H. T. (2018). Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. *Journal of Internet Services and Applications, 9*(1), 18.

Candelieri, A., & Archetti, F. Analyzing tweets to enable sustainable, multi-modal and personalized urban mobility: Approaches and results from the Italian project TAM-TAM. In 20th International Conference on Urban Transport and the Environment, UT 2014, Algarve, 2014 (Vol. 138, pp. 373-379). doi:https://doi.org/10.2495/UT140311.

Candelieri, A., & Archetti, F. Detecting events and sentiment on twitter for improving urban mobility. In 2nd International Workshop on Emotion and Sentiment in Social and Expressive Media, ESSEM 2015, 2015 (Vol. 1351, pp. 106–115)

Cao, J., Zeng, K., Wang, H., Cheng, J., Qiao, F., Wen, D., et al. (2014). Web-based traffic sentiment analysis: Methods and applications. *IEEE Transactions on Intelligent Transportation Systems, 15*(2), 844–853. https://doi.org/10.1109/TITS.2013.2291241

Carpenter, T., Golab, L., & Syed, S. J. Is the grass greener? Mining electric vehicle opinions. In 5th ACM International Conference on Future Energy Systems, e-Energy 2014, Cambridge, 2014 (pp. 241–252). doi:https://doi.org/10.1145/2602044.2602050.

Casas, I., & Delmelle, E. C. (2017). Tweeting about public transit — Gleaning public perceptions from a social media microblog. *Case Studies on Transport Policy, 5*(4), 634–642. https://doi.org/10.1016/j.cstp.2017.08.004

Chaniotakis, E., Antoniou, C., & Pereira, F. (2016). Mapping social media for transportation studies. *IEEE Intelligent Systems, 31*(6), 64–70.

Chen, S., Huang, Y., & Huang, W. Big Data Analytics on Aviation Social Media: The Case of China Southern Airlines on Sina Weibo. In 2nd IEEE International Conference on Big Data Computing Service and Applications, BigDataService 2016, 2016 (pp. 152–155). doi:https://doi.org/10.1109/BigDataService.2016.51.

Chen, Y., Lv, Y., Wang, X., Li, L., & Wang, F. (2018). Detecting traffic information from social media texts with deep learning approaches. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2018.2871269

D'Andrea, E., Ducange, P., Lazzerini, B., & Marcelloni, F. (2015). Real-time detection of traffic from twitter stream analysis. *IEEE Transactions on Intelligent Transportation Systems, 16*(4), 2269–2283. https://doi.org/10.1109/TITS.2015.2404431

Daly, E. M., Lecue, F., & Bicer, V. Westland row why so slow? Fusing social media and linked data sources for understanding real-time traffic conditions. In 18th International Conference on Intelligent User Interfaces, IUI 2013, Santa Monica, CA, 2013 (pp. 203–212). doi:https://doi.org/10.1145/2449396.2449423.

Dutta Das, D., Sharma, S., Natani, S., Khare, N., & Singh, B. Sentimental Analysis for Airline Twitter data. In 14th International Conference on Science, Engineering and Technology, ICSET 2017, 2017 (4 ed., Vol. 263). doi:https://doi.org/10.1088/1757-899X/263/4/042067.

Fiarni, C., Maharani, H., & Irawan, E. Implementing rule-based and naive bayes algorithm on incremental sentiment analysis system for Indonesian online transportation services review. In 10th

International Conference on Information Technology and Electrical Engineering, ICITEE 2018, 2018 (pp. 597–602). doi:https://doi.org/10.1109/ICITEED.2018.8534912.

Fu, K., Lu, C. T., Nune, R., & Tao, J. X. Steds: Social Media Based Transportation Event Detection with Text Summarization. In 18th IEEE International Conference on Intelligent Transportation Systems, ITSC 2015, 2015 (Vol. 2015-October, pp. 1952–1957). doi:https://doi.org/10.1109/ITSC.2015.316.

Gal-Tzur, A., Grant-Muller, S. M., Kuflik, T., Minkov, E., Nocera, S., & Shoor, I. (2014). The potential of social media in delivering transport policy goals. *Transport Policy, 32*, 115–123. https://doi.org/10.1016/j.tranpol.2014.01.007

Gal-Tzur, A., Rechavi, A., Beimel, D., & Freund, S. (2018). An improved methodology for extracting information required for transport-related decisions from Q&A forums: A case study of TripAdvisor. *Travel Behaviour and Society, 10*, 1–9. https://doi.org/10.1016/j.tbs.2017.08.001

Gao, L., Yu, Y., & Liang, W. (2016). Public transit customer satisfaction dimensions discovery from online reviews. *Urban Rail Transit, 2*(3–4), 146–152. https://doi.org/10.1007/s40864-016-0042-0

Georgiou, T., Abbadi, A. E., Yan, X., & George, J. Mining complaints for traffic-jam estimation: A social sensor application. In 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 25–28 Aug. 2015 2015 (pp. 330–335). doi:https://doi.org/10.1145/2808797.2809404.

Giancristofaro, G. T., & Panangadan, A. Predicting Sentiment toward Transportation in Social Media using Visual and Textual Features. In 19th IEEE International Conference on Intelligent Transportation Systems, ITSC 2016, 2016 (pp. 2113–2118). doi:https://doi.org/10.1109/ITSC.2016.7795898.

Grant-Muller, S. M., Gal-Tzur, A., Minkov, E., Nocera, S., Kuflik, T., & Shoor, I. (2014). Enhancing transport data collection through social media sources: Methods, challenges and opportunities for textual data. *IET Intelligent Transport Systems, 9*(4), 407–417.

Gu, Y., Qian, Z. S., & Chen, F. (2016). From twitter to detector: Real-time traffic incident detection using social media data. *Transportation Research Part c: Emerging Technologies, 67*, 321–342.

Guerrero-Ibáñez, J., Zeadally, S., & Contreras-Castillo, J. (2018). Sensor technologies for intelligent transportation systems. *Sensors, 18*(4), 1212.

Gupta, N., Crosby, H., Purser, D., Javis, S., & Guo, W. Twitter usage across industry: A spatiotemporal analysis. In 4th IEEE International Conference on Big Data Computing Service and Applications, BigDataService 2018, 2018 (pp. 64–71). doi:https://doi.org/10.1109/BigDataService.2018.00018.

Haghighi, N. N., Liu, X. C., Wei, R., Li, W., & Shao, H. (2018). Using twitter data for transit performance assessment: A framework for evaluating transit riders' opinions about quality of service. *Public Transport, 10*(2), 363–377. https://doi.org/10.1007/s12469-018-0184-4

Hoang, T., Cher, P. H., Prasetyo, P. K., & Lim, E. P. Crowdsensing and analyzing micro-event tweets for public transportation insights. In 4th IEEE International Conference on Big Data, Big Data 2016, 2016 (pp. 2157–2166). doi:https://doi.org/10.1109/BigData.2016.7840845.

Hosseini, M., El-Diraby, T., & Shalaby, A. (2018). Supporting sustainable system adoption: Socio-semantic analysis of transit rider debates on social media. [Article]. *Sustainable Cities and Society, 38*, 123–136. https://doi.org/10.1016/j.scs.2017.12.025

Hu, M., & Liu, B. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004 (pp. 168–177): ACM.

Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S., & Kitsuregawa, M. (2016). Visual exploration of changes in passenger flows and tweets on mega-city metro network. *IEEE Transactions on Big Data, 2*(1), 85–99. https://doi.org/10.1109/TBDATA.2016.2546301

Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (2019). Predicting elections from social media: A three-country, three-method comparative study. *Asian Journal of Communication, 29*(3), 252–273.

Kaur, W., & Balakrishnan, V. (2018). Improving sentiment scoring mechanism: A case study on airline services. *Industrial Management and Data Systems, 118*(8), 1578–1596. https://doi.org/10.1108/IMDS-07-2017-0300

Kim, K., Park, O. J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change, 123*, 362–369. https://doi.org/10.1016/j.techfore.2017.01.001

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and Software Technology, 51*(1), 7–15.

Kovács-Győri, A., Ristea, A., Havas, C., Resch, B., & Cabrera-Barona, P. (2018). #London2012: Towards citizen-contributed urban planning through sentiment analysis of twitter data. *Urban Planning, 3*(1), 75–99. https://doi.org/10.17645/up.v3i1.1287

Kuflik, T., Minkov, E., Nocera, S., Grant-Muller, S., Gal-Tzur, A., & Shoor, I. (2017). Automating a framework to extract and analyse transport related social media content: The potential and the challenges. *Transportation Research Part c: Emerging Technologies, 77*, 275–291. https://doi.org/10.1016/j.trc.2017.02.003

Kulkarni, G., Abellera, L., & Panangadan, A. Unsupervised classification of online community input to advance transportation services. In 8th IEEE Annual Computing and Communication Workshop and Conference, CCWC 2018, 2018 (Vol. 2018-January, pp. 261–267). doi:https://doi.org/10.1109/CCWC.2018.8301704.

Kumar, A., Jiang, M., & Fang, Y. Where not to go? Detecting road hazards using Twitter. In 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2014, Gold Coast, QLD, 2014 (pp. 1223–1226). doi:https://doi.org/10.1145/2600428.2609550.

Lacic, E., Kowald, D., & Lex, E. High enough? Explaining and predicting traveler satisfaction using airline reviews. In 27th ACM Conference on Hypertext and Social Media, HT 2016, 2016 (pp. 249–254). doi:https://doi.org/10.1145/2914586.2914629.

Lee, A. S. H., Yusoff, Z., Zainol, Z., & Pillai, V. (2018). Know your hotels well! An online review analysis using text analytics. *International Journal of Engineering & Technology, 7*(4.31), 341–437.

Lee, K., & Yu, C. (2018). Assessment of airport service quality: A complementary approach to measure perceived service quality based on Google reviews. *Journal of Air Transport Management, 71*, 28–44. https://doi.org/10.1016/j.jairtraman.2018.05.004

Liau, B. Y., & Tan, P. P. (2014). Gaining customer knowledge in low cost airlines through text mining. *Industrial Management and Data Systems, 114*(9), 1344–1359. https://doi.org/10.1108/IMDS-07-2014-0225

Liu, S., Tian, Y., Feng, Y., & Zhuang, Y. J. B. D. X. X. B. (2018). Comparison of tourist thematic sentiment analysis methods based on weibo data. *Beijing Da Xue Xue Bao, 54*(4), 687–692.

Liyang, H., Panangadan, A., & Abellera, L. V. Understanding public sentiment toward I-710 Corridor Project from social media based on Natural Language processing. In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 1–4 Nov. 2016 2016 (pp. 2107–2112). doi:https://doi.org/10.1109/ITSC.2016.7795897.

Lu, Z., Du, R., Dunham-Jones, E., Park, H., & Crittenden, J. (2017). Data-enabled public preferences inform integration of autonomous vehicles with transit-oriented development in Atlanta. *Cities, 63*, 118–127. https://doi.org/10.1016/j.cities.2017.01.004

Luckner, M., Kobojek, P., & Zawistowski, P. Public transport stops state detection and propagation warsaw use case. In 6th International Conference on Smart Cities and Green ICT Systems, SMARTGREENS 2017, 2017 (pp. 235–241)

Lv, Y., Chen, Y., Zhang, X., Duan, Y., & Li, N. L. (2017). Social media based transportation research: The state of the work and the networking. *IEEE/CAA Journal of Automatica Sinica, 4*(1), 19–26.

Mostafa, M. M. (2013). An emotional polarity analysis of consumers' airline service tweets. *Social Network Analysis and Mining, 3*(3), 635–649. https://doi.org/10.1007/s13278-013-0111-2

Musaev, A., Jiang, Z., Jones, S., Sheinidashtegol, P., & Dzhumaliev, M. (2018). Detection of damage and failure events of road infrastructure using social media. 25th International Conference on Web Services, ICWS 2018 Held as Part of the Services Conference Federation, SCF 2018 (Vol. 10966 LNCS, pp. 134–148).

Nikolaidou, A., & Papaioannou, P. (2018). Utilizing social media in transport planning and public transit quality: Survey of literature. *Journal of Transportation Engineering, Part a: Systems, 144*(4), 04018007.

O'Regan, G. (2018). The smartphone and social media. *World of Computing* (pp. 257–265). Springer.

Osborne, M., Moran, S., McCreadie, R., Von Lunen, A., Sykora, M., Cano, E., ... & O'Brien, A. (2014, June). Real-time detection, tracking, and monitoring of automatically discovered events in social media. In Proceedings of 52nd annual meeting of the association for computational linguistics: System demonstrations (pp. 37–42).

Patel, D. J., John, S. V., & Kaliangra, F. Managing traffic flow based on predictive data analysis. In Proceedings of International Conference on Advances in Computing, 2013 (pp. 1069–1074): Springer.

Pereira, C. K., Campos, F., Ströele, V., David, J. M. N., & Braga, R. (2018). BROAD-RSI–educational recommender system using social networks interactions and linked data. *Journal of Internet Services and Applications, 9*(1), 7.

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. Systematic mapping studies in software engineering. In Ease, 2008 (Vol. 8, pp. 68-77)

Pournarakis, D. E., Sotiropoulos, D. N., & Giaglis, G. M. (2017). A computational model for mining consumer perceptions in social media. *Decision Support Systems, 93*, 98–110. https://doi.org/10.1016/j.dss.2016.09.018

Rahman, S. S., Easton, J. M., & Roberts, C. Mining open and crowdsourced data to improve situational awareness for railway. In 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 25–28 Aug. 2015 2015 (pp. 1240–1243). doi:https://doi.org/10.1145/2808797.2809369.

Rane, A., & Kumar, A. Sentiment Classification System of Twitter Data for US Airline Service Analysis. In 42nd IEEE Computer Software and Applications Conference, COMPSAC 2018, 2018 (Vol. 1, pp. 769–773). doi:https://doi.org/10.1109/COMPSAC.2018.00114.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part c: Emerging Technologies, 75*, 197–211.

Rybarczyk, G., Banerjee, S., Starking-Szymanski, M. D., & Shaker, R. R. (2018). Travel and us: the impact of mode share on sentiment using geo-social media and GIS. *Journal of Location Based Services, 12*(1), 40–62. https://doi.org/10.1080/17489725.2018.1468039

Salas, A., Georgakis, P., Nwagboso, C., Ammari, A., & Petalas, I. Traffic event detection framework using social media. In 2017 IEEE International Conference on Smart Grid and Smart Cities, ICSGSC 2017, 2017 (pp. 303–307). doi:https://doi.org/10.1109/ICSGSC.2017.8038595.

Salas, A., Georgakis, P., & Petalas, Y. Incident detection using data from social media. In 20th IEEE International Conference on Intelligent Transportation Systems, ITSC 2017, 2018 (Vol. 2018-March, pp. 751–755). doi:https://doi.org/10.1109/ITSC.2017.8317967.

Saldana-Perez, A. M. M., Moreno-Ibarra, M., & Tores-Ruiz, M. Classification of traffic related short texts to analyse road problems in urban areas. In 2nd International Conference on Smart Data and Smart Cities, UDMS 2017, 2017 (4W3 ed., Vol. 42, pp. 91-97). Doi: 10.5194/isprs-archives-XLII-4-W3-91-2017.

Samonte, M. J. C., Dollete, C. J. T., Capanas, P. M. M., Flores, M. L. C., & Soriano, C. B. (2018). Sentence-Level Sarcasm Detection in English and Filipino Tweets. Paper presented at the Proceedings of the 4th International Conference on Industrial and Business Engineering, Macau, Macao,

Saragih, M. H., & Girsang, A. S. Sentiment analysis of customer engagement on social media in transport online. In 2017 International Conference on Sustainable Information Engineering and Technology, SIET 2017, 2018 (Vol. 2018-January, pp. 24–29). doi:https://doi.org/10.1109/SIET.2017.8304103.

Schulz, R., Wyeth, G., & Wiles, J. (2010). Language change across generations for robots using cognitive maps. *ALIFE* (pp. 581–588). Citeseer.

Sdoukopoulos, A., Nikolaidou, A., Pitsiava-Latinopoulou, M., & Papaioannou, P. (2018). Use of social media for assessing sustainable urban mobility indicators. *International Journal of Sustainable Development and Planning, 13*(2), 338–348. https://doi.org/10.2495/SDP-V13-N2-338-348

Septiana, I., Setiowati, Y., & Fariza, A. Road condition monitoring application based on social media with text mining system: Case Study: East Java. In 18th International Electronics Symposium, IES 2016, 2017 (pp. 148–153). doi:https://doi.org/10.1109/ELECSYM.2016.7860992.

Serna, A., & Gasparovic, S. TRANSPORT ANALYSIS APPROACH BASED on BIG DATA and TEXT MINING ANALYSIS from SOCIAL MEDIA. In 13th Conference on Transport Engineering, CIT 2018, 2018 (Vol. 33, pp. 291–298). doi:https://doi.org/10.1016/j.trpro.2018.10.105.

Serna, A., Gerrikagoitia, J. K., Bernabé, U., & Ruiz, T. Sustainability analysis on Urban Mobility based on Social Media content. In Transportation Research Procedia, 2017 (Vol. 24, pp. 1–8). doi:https://doi.org/10.1016/j.trpro.2017.05.059.

Sharma, K., & Mediratta, P. (2002). Importance of keywords for retrieval of relevant articles in medline search. *Indian Journal of Pharmacology, 34*(5), 369.

Sinha, M., Varma, P., & Mukherjee, T. Web and social media analytics towards enhancing urban transportations: A case for Bangalore. In 2nd ACM SIGMOD Workshop on Network Data Analytics, NDA 2017, 2017. doi:https://doi.org/10.1145/3068943.3068950.

Sloan, L., & Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLoS ONE, 10*(11), e0142209. https://doi.org/10.1371/journal.pone.0142209.

Sternberg, F., Hedegaard Pedersen, K., Ryelund, N. K., Mukkamala, R. R., & Vatrapu, R. Analysing Customer Engagement of Turkish Airlines Using Big Social Data. In 7th IEEE International Congress on Big Data, BigData Congress 2018, 2018 (pp. 74–81). doi:https://doi.org/10.1109/BigDataCongress.2018.00017.

Suma, S., Mehmood, R., Albugami, N., Katib, I., & Albeshri, A. Enabling Next Generation Logistics and Planning for Smarter Societies. In 8th International Conference on Ambient Systems, Networks and

Technologies, ANT 2017 and 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 2017 (Vol. 109, pp. 1122–1127). doi:https://doi.org/10.1016/j.procs.2017.05.440.

Thelwall, M. (2017). Tensistrength: Stress and relaxation magnitude detection for social media texts. *Information Processing and Management, 53*(1), 106–121. https://doi.org/10.1016/j.ipm.2016.06.009

Tse, R., Xiao, Y., Pau, G., Fdida, S., Roccetti, M., & Marfia, G. (2016). Sensing pollution on online social networks: A transportation perspective. *Mobile Networks and Applications, 21*(4), 688–707. https://doi.org/10.1007/s11036-016-0725-5

Ulloa, D., Saleiro, P., Rossetti, R. J. F., & Silva, E. R. Mining social media for open innovation in transportation systems. In 19th IEEE International Conference on Intelligent Transportation Systems, ITSC 2016, 2016 (pp. 169–174). doi:https://doi.org/10.1109/ITSC.2016.7795549.

Wang, C., Pan, X., & Wang, Y. Social networks and railway passenger capacity: An empirical study based on text mining and deep learning. In 4th ACM SIGSPATIAL International Workshop on Safety and Resilience, EM-GIS 2018, held with the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL 2018, 2018. doi:https://doi.org/10.1145/3284103.3284125.

Wang, D., Al-Rubaie, A., Clarke, S. S., & Davies, J. (2017). Real-time traffic event detection from social media. *Acm Transactions on Internet Technology, 18*(1), 1–23. https://doi.org/10.1145/3122982

Wang, D., Al-Rubaie, A., Davies, J., & Clarke, S. S. Real time road traffic monitoring alert based on incremental learning from tweets. In 2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS), 2014 (pp. 50–57): IEEE.

Wayasti, R. A., Surjandari, I., & Zulkarnain. Mining customer opinion for topic modeling purpose: Case study of ride-hailing service provider. In 6th International Conference on Information and Communication Technology, ICoICT 2018, 2018 (pp. 305–309). doi:https://doi.org/10.1109/ICoICT.2018.8528751.

Windasari, I. P., Uzzi, F. N., & Satoto, K. I. Sentiment analysis on Twitter posts: An analysis of positive or negative opinion on GoJek. In 4th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2017, 2018 (Vol. 2018-January, pp. 266–269). doi:https://doi.org/10.1109/ICITACEE.2017.8257715.

Yang, J., & Anwar, A. M. Social Media Analysis on Evaluating Organisational Performance: A Railway Service Management Context. In 14th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 14th IEEE International Conference on Pervasive Intelligence and Computing, PICom 2016, 2nd IEEE International Conference on Big Data Intelligence and Computing, DataCom 2016 and 2016 IEEE Cyber Science and Technology Congress, CyberSciTech 2016, DASC-PICom-DataCom-CyberSciTech 2016, 2016 (pp. 835–841). doi:https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.143.

Zakari, A., Lee, S. P., Alam, K. A., & Ahmad, R. (2018). Software fault localisation: A systematic mapping study. *IET Software, 13*(1), 60–74.

Zhang, B., Kotkov, D., Veijalainen, J., & Semenov, A. (2016). Online stakeholder interaction of some airlines in the light of situational crisis communication theory. 15th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2016 (Vol. 9844 LNCS, pp. 183–192).

Zhang, L., Sun, Y., & Luo, T. A framework for evaluating customer satisfaction. In 10th International Conference on Software, Knowledge, Information Management and Applications, SKIMA 2016, 2017 (pp. 448–453). doi:https://doi.org/10.1109/SKIMA.2016.7916264.

Zhang, X., Dong, X., Wu, J., Cao, Z., & Lyu, C. (2017). Fault activity aware service delivery in wireless sensor networks for smart cities. *Wireless Communications and Mobile Computing, 2017*, 1–22.

Zhang, X., Niu, S., Zhang, D., Wang, G. A., & Fan, W. (2015). Predicting vehicle recalls with user-generated contents: A text mining approach. 10th Pacific Asia Workshop on Intelligence and Security Informatics, PAISI 2015 in Conjunction with Pacific-Asia Conference on Knowledge Discovery and Data Mining , PAKDD 2015 (Vol. 9074, pp. 41–50).

Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., & Philip, S.Y.J.K.-B.S. (2018). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems, 143*, 236–247.

Zhang, Z., Chen, S., Yuan, S., & Zhang, J. (2018). A combinational classification for the customers of airline platform based on text mining. 4th International Conference on Fuzzy Systems and Data Mining, FSDM 2018 (Vol. 309, pp. 302–312).

Zhao, S., Gao, Y., Ding, G., & Chua, T. S. (2017). Real-time multimedia social event detection in microblog. *IEEE Transactions on Cybernetics, 48*(11), 3218–3231.