Estimate Sentiment of Crowds from Social Media during City Events

Gong, Vincent X.; Daamen, Winnie; Bozzon, Alessandro; Hoogendoorn, Serge P.

DOI
10.1177/0361198119846461

Publication date
2019

Document Version
Final published version

Published in
Transportation Research Record

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–15


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0361198119846461
journals.sagepub.com/home/trr

Estimate Sentiment of Crowds from Social Media during City Events

Vincent X. Gong[1,2], Winnie Daamen[1], Alessandro Bozzon[2], and Serge P. Hoogendoorn[1]

Abstract

City events are being organized more frequently, and with larger crowds, in urban areas. There is an increased need for novel methods and tools that can provide information on the sentiments of crowds as an input for crowd management. Previous work has explored sentiment analysis and a large number of methods have been proposed relating to various contexts. None of them, however, aimed at deriving the sentiments of crowds using social media in city events, and no existing event-based dataset is available for such studies. This paper investigates how social media can be used to estimate the sentiments of crowds in city events. First, some lexicon-based and machine learning-based methods were selected to perform sentiment analyses, then an event-based sentiment annotated dataset was constructed. The performance of the selected methods was trained and tested in an experiment using common and event-based datasets. Results show that the machine learning method LinearSVC achieves the lowest estimation error for sentiment analysis on social media in city events. The proposed event-based dataset is essential for training methods to reduce estimation error in such contexts.

As cities compete for global attractiveness and community quality, city-scale public events become more and more popular to boost tourism and promote economic growth. Thematic exhibitions, sports competitions, and national celebrations are instances of city events that take place in urban areas, and may attract a large amount of people during a short time period. The scale and intensity of these events require technical solutions that support stakeholders (e.g., event organizers and public and safety authorities) to manage the crowd.

During such events, the crowd is managed by public authorities to reduce the risk of incidents as a result of internal and external threats. This is usually achieved by exerting predefined measures based on qualitative interpretations of the crowd by stewards, police officers, or event organization employees.

As the efficiency and effectiveness of crowd management measures depend on pedestrian behavior (*1, 2*), it is beneficial for stakeholders to obtain information about the behavior of the crowd. The sentiment of people in the crowd is one of the factors affecting crowd behavior (*3*). Together with other information such as crowd density and demographics, it may help crowd managers estimate and predict (negative) behaviors that can be inferred from the sentiment of people in the crowd, such as risky behaviors. Therefore, deriving the sentiment of people in the crowd could be valuable to crowd management.

The sentiment of crowds is difficult to acquire, however. In conventional approaches this information is captured manually by stewards or staff members (*4*), a practice that is costly and subject to bias. Traditional crowd observation techniques are based on sensors (e.g., counting systems, GPS trackers, and Wi-Fi sensors) which only provide spatio-temporal information. These solutions do not provide sentiment values. Although crowd sentiment could be extracted from image or video clips provided by cameras through image recognition techniques (*5, 6*), accessing the images or video recordings of a public area is computationally intensive, and often restricted because of privacy issues.

The advance of web-based technologies provides new data sources which could be applied to understand and analyze pedestrian behavior (*7–10*). Several social media networks, such as Twitter and Instagram, are widely used. Time-stamped social media posts, such as text content, are often geo-tagged. More importantly, these posts intrinsically embrace rich semantic information which

[1]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[2]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

Corresponding Author:
Address correspondence to Vincent X. Gong: x.gong-1@tudelft.nl

could be employed for deriving sentiments in a crowd. Therefore, social media data can be used to derive the sentiments of crowds during events.

A large number of works studied sentiment estimation of crowds using social media data. Jiang et al. (*11*) introduced a method to estimate sentiment of tweets considering multiple strategies and including context information (i.e., related tweets). Zhang et al. (*12*) demonstrated a machine-learning method incorporating syntactic and context information from social media to estimate users' sentiments. Ortigosa et al. (*13*) presented a method to estimate sentiment of users based on their Facebook texts which integrated lexicon-based, machine-learning techniques.

Yu et al. (*14*) applied sentiment analysis to the financial sector, deriving the sentiment of traders from social networks, and exploring its relationship with short-term stock market performance. Surveys about sentiment analysis (*15, 16*) reviewed more than 20 methods and 30 works on estimating sentiment from social media. The methods can be categorized into three types: lexicon-based methods, machine learning (ML)-based methods, and hybrid methods. Lexicon-based methods assign each consecutive combination of words of a text a sentiment score according to a dictionary and calculate the weighted average sentiment score. ML-based methods train the model with a sentiment annotated dataset and estimate the sentiment of a test dataset through the model. Hybrid methods are a combination of lexicon- and ML-based methods. With respect to the context, sentiment analysis in the context of city events for crowd management differs from other contexts (e.g., E-learning, marketing, stock market prediction) in a set of characteristics, such as the specific topic of the event, its location, popularity, and time of occurrence. Consequently, sentiment analysis methods suitable for other contexts may differ from methods fit for the context of city events. While showing the utility of social media data in sentiment analysis studies, no previous work has aimed at deriving the sentiments of crowds in the context of city events. Moreover, the sentiment annotated dataset for a specific context is significant in sentiment analysis. It can be used for evaluating sentiment estimation results from various methods. It can also be used for ML-based or hybrid methods to train their models for sentiment estimation. However, none of the previous works proposed sentiment annotated datasets in such a context. There is a lack of an in-depth understanding of which methods are most effective in this context, and whether their performance will be affected by the diversity of the events or the urban areas in which they take place.

These research gaps lead to the following research question: Which methods are suitable to derive sentiments of crowds from social media texts in city events?

To answer this research question, a number of methods were selected. To compare and assess their performance in the context of city events, an event-based sentiment annotated dataset was required as ground truth. As no annotated event-based datasets existed, the authors constructed one. Using this dataset, the authors tested the performance of candidate methods and selected the most promising method.

The next section presents a literature review, followed by the research methodology to examine the performance of candidate methods. In the fourth section, the methods for comparison are selected, followed by a description of the data collection. The fifth section introduces the experimental setting, followed by the findings and analysis and discussion of the results. Conclusions and proposals for future work are presented at the end of the paper.

Literature Review

The present work compares the sentiment analysis performance of various methods and proposes an event-based sentiment dataset. This section briefly reviews previous works about comparison of sentiment analysis methods and the proposed sentiment datasets.

Previous works have performed experiments in certain contexts or for certain purposes, such as for document classification, e-learning, and brand marketing. They select a set of methods for comparison, use datasets to train their model, and evaluate the results. Therefore, three elements are involved: the context for comparing methods, the selected methods for comparison, and the datasets for training and testing. The following literature review is structured with regard to these three elements.

Sentiment analysis has been applied in various contexts. Pang et al. (*15*) investigated a set of methods to estimate sentiment for classifying documents. They compared the sentiment analysis performance of several ML methods based on public reviews collected from the internet, such as movie reviews from IMDb (an online database of information related to films, television programs, home videos and video games, and internet streams and fan reviews and ratings: https://www.imdb.com/). The methods selected for comparison were Naive Bayes (NB), support vector machines (SVMs), and maximum entropy (ME). The results showed that SVM outperformed other methods in their experiment. Boiy and Moens (*17*) applied sentiment analysis for opinion mining on multilingual web texts. They analyzed the performance of ML methods including NB, SVM, and ME to estimate sentiment in public reviews about cars and movies. Their findings showed that SVM outperformed other methods.

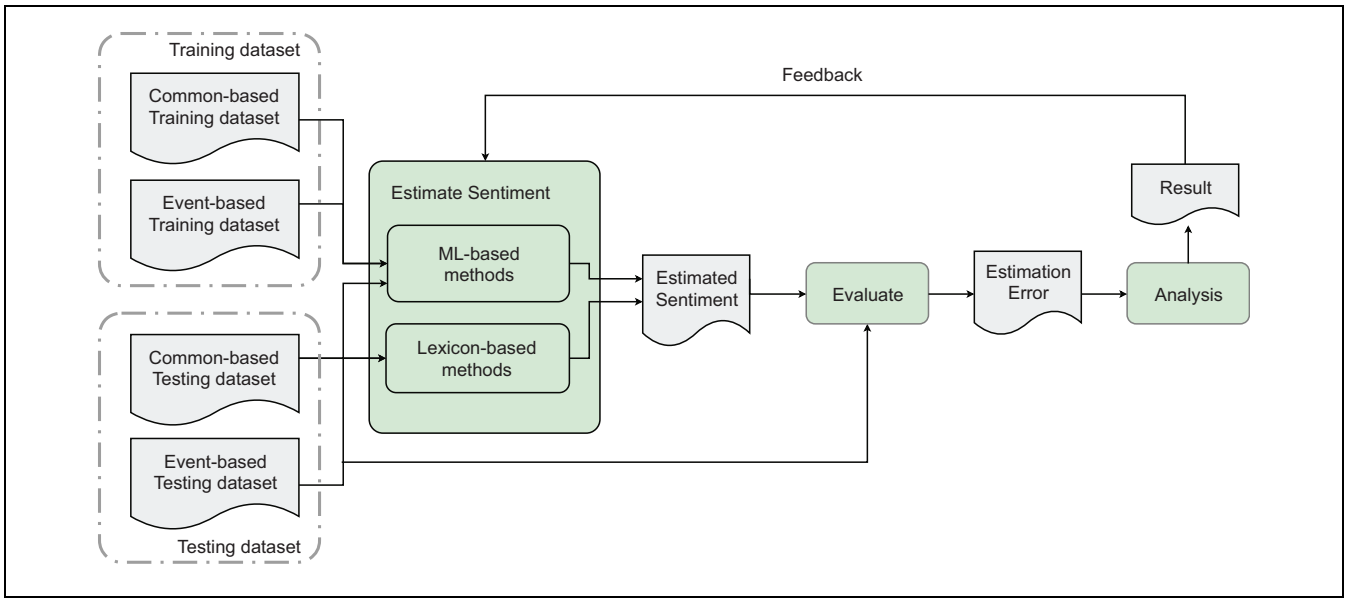

Figure 1. The process of investigating sentiment estimation performance of selected methods on social media text in city events. The green symbols denote three major steps: estimate the sentiment using different methods with different datasets; evaluate the estimation error per method; and analyze the estimation error between methods. The gray symbols denote documents input and output.

Li and Li (*18*) estimated sentiment on social media for marketing purposes. They derived market intelligence through sentiment analysis based on product reviews on Twitter, such as reviews for Microsoft and Sony products, iPhone, iPad, and Macbook. They compared the performance of NB and SVM with various settings and found that SVM achieved the best performance in their experiment. Ortigosa et al. (*13*) analyzed sentiment in the context of e-learning based on social media data. A set of ML methods including a hybrid method were compared for sentiment analysis of text from Facebook. In contrast to other works, they found that NB outperformed other methods, including SVM, in this context. Bravo-Marquez et al. (*19*) studied sentiment analysis on big social data. They compared sentiment through various methods with different configurations based on a large volume of non-topic specific posts on Twitter ("tweets"). They also found NB to perform better than SVM in their experiment.

This review of the literature shows that NB and SVM are the most popular ML methods for sentiment analysis (*13, 15, 17–21*). Public reviews and social media data are widely used for deriving sentiment. Some investigations (*13, 15, 17, 18*) have also been performed with data related to certain topics, such as reviews of movies or cars, as well as brands such as Microsoft, Sony, and Google, while other studies (*19–21*) were performed with common-based datasets. None of these works, however, was performed in the context of a city event with respect to crowd management, and none of them investigated the impact of using different datasets, that is, common-based or a certain topic based, on the performance of sentiment analysis.

Regarding the construction of datasets, Kouloumpis et al. (*22*) published a dataset including 222,570 sentences annotated with three sentiment categories: positive, neutral, and negative. The sentences were collected from Twitter with no specific topic. Costa et al. (*23*) published a comprehensive dataset containing 400 deceptive and 400 truthful reviews in positive and negative categories. These datasets have been widely used in several sentiment analysis works (*24–28*). There are also datasets proposed related to certain topics. For instance, Hu and Liu (*29*) presented a dataset with 6,800 opinion words on 10 different products. Cruz et al. (*30*) proposed a sentiment annotated database of reviews with different topics: 587 reviews of headphones, 988 reviews of hotels, and 972 reviews of cars. Similarly Blitzer et al. (*31*) proposed a sentiment annotated dataset including Amazon reviews in four domains: books, DVDs, electronics, and kitchen appliances. However, there is no dataset proposed for sentiment analysis in the context of city events.

Research Approach

This section elaborates the research approach for testing the sentiment estimation performance of candidate methods in the context of city events using social media.

The research approach is illustrated in Figure 1. It consists of three major steps: (i) estimate the sentiment using different methods and different datasets, (ii) evaluate the estimation error per method, and (iii) compare

the estimation error between methods. It involves two different types of datasets: common-based datasets and an event-based dataset. The common-based datasets in this research cover a wide variety of situations and have no domain knowledge or context information about city events. Moreover, these datasets are well known in the research area of sentiment analysis. Event-based datasets refers to datasets of social media posts collected during events and annotated with sentiments. The event-based dataset used was generated in the course of this research. In the first step, lexicon-based methods estimate the sentiment of each text in both common and event-based datasets. The ML-based methods train their models using part of the common and event-based datasets. Then both trained models estimate the sentiments of the remaining texts in the common and event-based datasets, respectively. In this step, both methods yield estimated sentiment. In the second step, to verify whether the estimated sentiment is correct, the estimation result is compared with the sentiment ground truth of the test dataset using the metrics introduced in the next paragraph. N-round testing is then performed to reduce the random error. In each round, a subset is selected, and the sentiment of each text is identified and compared with the ground truth. Finally, the performance of the methods across different datasets is compared. The analysis results also provide feedback on the research methodology, such as how best to adjust the sample size of the training and testing dataset, select feasible candidate methods, and choose suitable comparison metrics.

Comparison Metrics

The sentiment estimation performance is assessed using the estimation error, which is calculated for each repetition of sentiment estimation. This estimation error per repetition E_i is calculated by the amount of false identifications M_i^{false} divided by the testing sample size M_i, see Equation 1. While running an N-round testing, the mean and standard deviation of the estimation error is calculated in each round.

$$E_i = \frac{M_i^{\text{false}}}{M_i} \tag{1}$$

Selection of Candidate Methods from Literature

This section presents the selection of a set of candidate methods to perform sentiment analysis on social media data. As indicated above, there is no existing literature comparing the performance of sentiment analysis methods using social media in the context of large-scale city events. Thus, the sentiment analysis methods reviewed were those applied to generic situations.

Deriving sentiments from social media text is not a novel problem. Many authors have discussed this topic and proposed methods to solve it (*16, 32, 33*). Ravi and Ravi (*16*) reviewed 161 studies of which about 30 discussed sentiment analysis on social media networks. As mentioned in the introductory section above, sentiment analysis methods can be categorized into three types: lexicon-based, ML-based, and hybrid methods. Hybrid methods are a combination of lexicon- and ML-based methods. The performance of hybrid methods is therefore influenced by the quality of the lexicon- and ML-based methods they combine. Understanding the performance of lexicon- and ML-based methods is necessary to investigate hybrid methods. Therefore, this research focuses on lexicon- and ML-based methods, and leaves hybrid methods for future work.

An overview of the methods selected for sentiment estimation on social media texts in the context of city events is shown in Table 1. More details on the selected methods are given below.

Lexicon-Based Methods

Lexicon-based, or dictionary based, approaches are widely applied in the field of sentiment analysis (*16, 34*). Given a text from a social media post, lexicon-based methods assign each n-gram (i.e., consecutive combination of words) a sentiment score according to its attached dictionary and calculate the weighted average sentiment score as a performance indicator after filtering out stop words and reducing other noises. More than 41 studies explore lexicon-based methods in sentiment analysis (*16*). Among these, SentiStrength and SentiWordNet are two popular lexicon-based methods used for deriving sentiment from social media data.

SentiStrength. SentiStrength was created by identifying sentiments expressed in the texts on MySpace, a social media platform. It estimates the strength of negative, neutral, and positive sentiment in short texts. It was originally developed for the English language and optimized for short social media texts (*20*). SentiStrength reports three sentiment values with a range of strengths: -5 to -1 as negative, 0 as neutral, and 1 to 5 as positive. It has been applied and investigated in many papers in which it shows significant performance (*21, 35–37*).

SentiWordNet. SentiWordNet is a lexical resource for opinion mining. Instead of constructing its sentiment dictionary from a corpus (e.g., MySpace data) as SentiStrength does, it assigns to each syncset of WordNet one of three sentiment values: positive,

Table 1. Selected Methods for Deriving Sentiment of Crowd on Social Media in City Events

Category	Method		Description	Linear or nonlinear	References[a]
Lexicon-based	SentiStrength		Optimized for social media text	na	*20, 21, 35, 36, 37*
	SentiWordNet		Assigns to WordNet synset	na	*20, 25, 38, 39, 40*
ML-based	Naive Bayes (NB)		Bayes theorem	Linear	*13, 14, 19, 42, 43*
	SVM	SGDClassifier	Fitted with stochastic gradient descent learning	Linear	*44, 45, 46*
		LinearSVC	Linear support vector classification	Linear	*11, 42, 46, 47*
		NuSVC	Statistical, Nu-support vector classification	Nonlinear	*13, 19, 20, 43, 46*
		SVC	Statistical, C-support vector classification	Nonlinear	

Note: na = not applicable; ML-based = machine learning-based; SVM = support vector machines; SVC = support vector classifier.
[a]For reference numbers, see References section.

Table 2. Sentiment Estimation Output Schema Transformation Rules

Category	Name	Output scheme	Convert rule	Unified schema
Lexicon-based	SentiStrength	–5 to −1 as negative, 0 as neutral, 1 to 5 as positive	[–5, −1] as −1, 0 as 0, [1, 5] as 1	−1 denotes negative, 0 denotes neutral, 1 denotes positive
	SentiWordNet	negative, neutral, positive	negative as −1, neutral as 0, positive as 1	
ML-based	NB SGDClassifier LinearSVC NuSVC SVC	−1, 0, 1 or −1, 1	na	

Note: na = not applicable; ML-based = machine learning-based; NB = naive Bayes; SVM = support vector machines; SVC = support vector classifier.

negative, or objective (*38*). It is widely used in estimating sentiment from social media networks (*20, 25, 39, 40*).

ML-Based Methods

ML-based methods train the model with a sentiment annotated dataset and estimate the sentiment of a test dataset through the model. Numerous ML-based methods have been proposed and investigated in recent studies in various situations (*16*). Among these methods, NB and SVM are widely tested and outperform most other methods in deriving sentiments from social media texts (*15, 16, 33*). In the following these methods are described in more detail.

Naive Bayes (NB). NB is a supervised linear ML algorithm which is popular for classifying text. It is a simple probabilistic classifier based on applying Bayes's theorem (*41*). It is widely used to estimate sentiments from social media texts (*13, 19, 42*). Although its mechanism is fairly straightforward, it often performs as well as much more complicated solutions (*14, 43*).

Support Vector Machines (SVMs). SVMs are a family of supervised learning models used for linear and nonlinear classification analysis. SVMs are widely used in text categorization for sentiment analysis (*15, 16*). In this research, the four most popular SVM models are tested, namely: stochastic gradient descent classifier fitted SVM (SGDClassifier or SGDC), linear support vector classifier (LinearSVC), Nu-support vector classifier (NuSVC), and support vector classifier (SVC). SGDClassifier is a linear SVM classifier fitted with stochastic gradient descent learning. LinearSVC is an implementation of SVC in case of a linear kernel. SVC and NuSVC apply the statistics of support vectors developed in the SVM algorithm. SVC and NuSVC are similar methods, but accept slightly different sets of parameters and have different mathematical formulations. These methods are explored in a large number of papers on deriving sentiment from social media (*11, 13, 19, 20, 42–47*).

Sentiment Estimation Result Scheme

This research aims to compare the performance of sentiment analysis methods. However, the various methods selected result in different sentiment schemes. For instance, the lexicon-based method SentiStrength outputs sentiment values as an integer between −5 and 5, while the SentiWordNet results in values of negative,

neutral, and positive. For other methods, the output schemes are listed in Table 2. To compare the performance of these selected methods, it is necessary to define a unified output scheme and map schemes of all the selected methods to the unified output scheme.

According to Table 2, there are two types of sentiment scheme: simplified or detailed. "Simplified" in this research refers to a sentiment scheme featuring only three categories: positive (1), neutral (0), and negative (–1). A detailed sentiment scheme, in contrast, has more sentiment categories, for example: extremely negative, very negative, negative, slightly negative, neutral, slightly positive, positive, very positive, extremely positive.

For lexicon-based methods, SentiStrength supports a detailed sentiment scheme, while SentiWordNet results in the simplified scheme. For ML-based methods, the supported sentiment scheme depends on the training data. Namely, if the training dataset is annotated with a detailed sentiment scheme, the ML-based methods trained with such a dataset also yield sentiment scores in the same scheme.

When constructing such dataset, however, the agreement reached on a sentiment category from a detailed scheme, for example, "extremely negative," is less than on a category from the simplified scheme, for example, "negative." Moreover, subjective errors introduced by human agents in the annotation process is also increased when using the detailed scheme. Thus, a dataset annotated with a detailed sentiment scheme is difficult to construct, less reliable, and therefore more rare. Most of the existing sentiment datasets are annotated with a simplified scheme, that is: negative, neutral, and positive. ML-based methods trained with such a dataset also result in a simplified sentiment scheme.

With regard to the impact of a simplified sentiment scheme on the estimation error of the models, compared with a detailed sentiment scheme, a simplified one indeed may lose the detailed sentiment strength information, but it still reports the same sentiment polarity; for example, in a detailed scheme, either "very positive" or "slightly positive" will be reported as "positive" in a simplified scheme.

As the simplified sentiment scheme is widely supported by both lexicon- and ML-based methods, the simplified sentiment score is applied in this research. The following three sentiment values are assigned: -1, 0, or 1, denoting negative, neutral, and positive, respectively. The mapping of all the selected methods is shown in Table 2.

Data Collection

Investigating the performance of candidate methods in the context of city events requires ground truth data, both for testing purposes and, in case of ML-based methods, to train their model. This research, required both common-based and event-based sentiment annotated social media data. Annotation in this respect means that, for each text, its sentiment is known. Common-based datasets cover a wide variety of situations and have no domain knowledge or context information about city events, while event-based datasets focus on posts which have been collected during events. As indicated above, annotated common-based datasets are available from previous research, but annotated event-based datasets are not yet available. To fill this research gap, the authors constructed such an event-based dataset.

Both the common and event-based sentiment datasets were annotated with sentiment polarities: Positive, Neutral, and Negative. Activities in city events, for example, celebrations and riots, tend to stimulate attendees' sentiments. The sentiments of crowds in the context of city events are stronger than in a normal context, meaning that they generate more extreme (positive or negative) expressions than neutral ones. Thus, it is valuable to explore the distinction of sentiment estimation with and without neutral polarity. This research considers sentiment polarity in two sets: one consists of Positive and Negative (PN) and the other Positive, Neutral, and Negative (PNN). Sentiment analysis with and without other individual polarities will be kept for future research.

The following subsections describe the selection of the common-based datasets and the construction of an annotated event-based dataset.

Common-Based Dataset

There is no official definition for a common-based dataset. In this research, it refers to datasets which cover a wide variety of situations, and contain no domain knowledge or context information about city events. Several papers have proposed sentiment annotated datasets consisting of texts collected from social media. The most comprehensive review (*16*) listed 32 public datasets used for sentiment analysis, six of which are social media datasets.

These datasets vary in relation to topic, sentiment polarity, and annotation approaches. Social media posts contained in these datasets may cover diverse topics, such as digital brands, sports, and technology. With regard to sentiment polarity, some datasets contain posts with Positive, Neutral, and Negative polarities, while others only have Positive and Negative posts. There are two major annotation approaches: by the researchers themselves or through crowd-sourcing. For the common-based ground truth in this research, the authors chose two social media datasets based on their large amount of

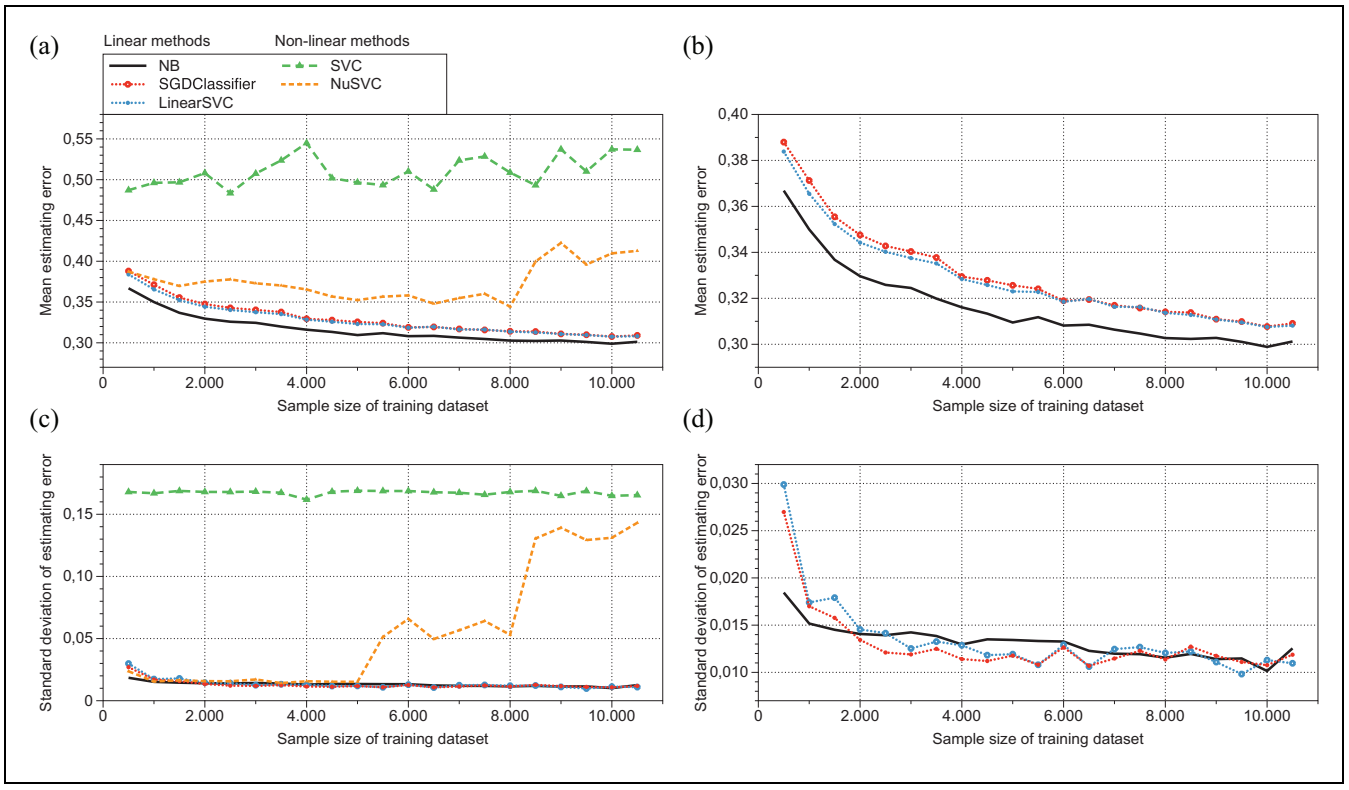

Figure 2. Mean and standard deviation of estimation error when increasing the training data sample size for ML-based methods. (*a*) Mean estimating error; (*b*) mean estimating error-linear methods; (*c*) standard deviation estimating error; and (*d*) standard deviation estimating error–linear methods. (*b*) and (*d*) zoom in linear ML-based methods to present the variation pattern of linear methods more clearly. The estimation error of linear ML-based methods stabilizes when the training sample size is higher than 6,000.

texts, the fact that they are widely used in other research, and their diverse sentiment polarities and topics. The first one is the dataset for the University of Michigan Sentiment Analysis competition, which consists of more than 1.5 million social media posts annotated as Positive or Negative. It is applied several times as the ground truth for this competition. The second dataset is an extended version of Niek Sanders's sentiment dataset series which is widely used in sentiment analysis studies (*48–51*). This dataset contains more than 55,000 social media posts, each of which is annotated with Positive, Neutral, or Negative. The posts in both datasets cover random topics.

Event-Based Dataset

Event-based datasets in this research refers to sentiment annotated datasets consisting of social media posts posted during both city-scale events and local events. They should be sufficiently large to be used as ground truth for testing candidate methods, but also serve as training data for ML-based methods. To construct such an event-based dataset, the first step was to estimate the required size of the dataset. The authors then collected social media posts and annotated them with sentiment scores. This process is elaborated upon in the following.

The size of an event-based dataset should meet two criteria. First, it should be sufficient as a training dataset for ML-based candidate methods to reach stabilized performance for sentiment analysis. Second, it should be as small as possible given the efforts and costs involved in performing the sentiment annotation. To estimate the sample size, a common-based dataset was used to investigate the estimation performance variance for different sample sizes using different ML-based methods, as shown in Figure 2. Figure 2*a* shows the variance of mean estimation error with respect to the size of the training sample. As expected, the mean error of linear ML-based methods decreases when the training sample is increased. However, the nonlinear ML-based methods show unexpected increases with increasing sample size, which may be caused by their nonlinear nature. To present the variation pattern of the linear method more clearly, an error is highlighted in Figure 2*b*. The figure shows that the mean estimation errors of linear ML-based methods decrease considerably when the training sample is less than 2,000 posts, after which the decrease in error becomes less

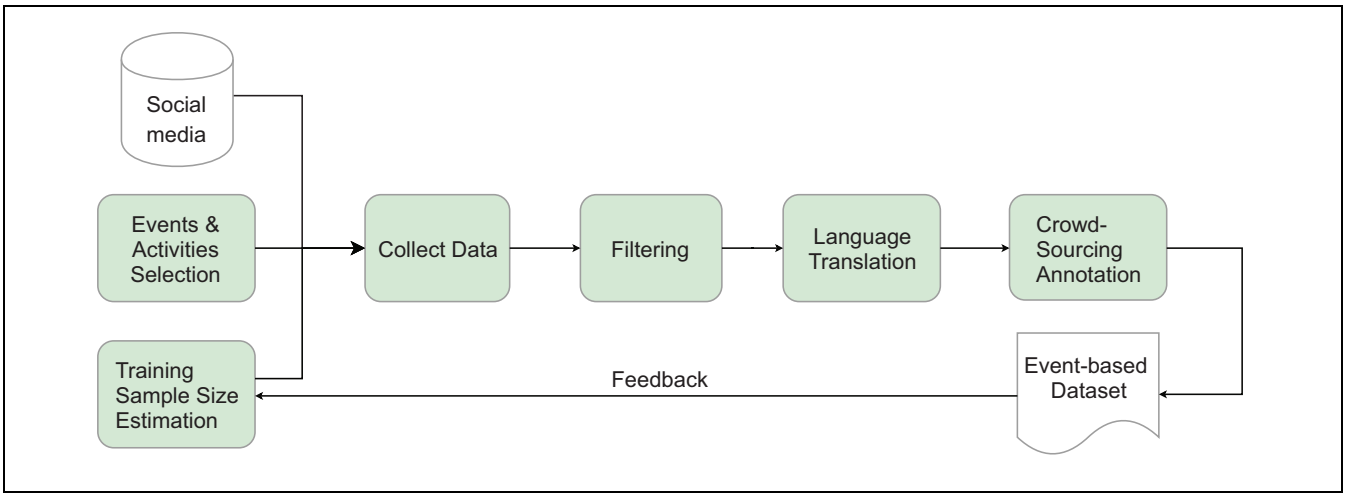

Figure 3. The process of constructing the annotated event-based social media sentiment dataset.

sharp and gradually flattens after a sample size of 6,000 posts. Taking a training sample size of 6,000, increasing the sample with another 2,000 posts only reduces the mean estimation error by less than 0.01. This also holds for the decrease of the standard deviation of estimation error, shown in Figure 2*c* and *d*. It can thus be concluded that the estimation error of linear ML-based methods stabilizes when the training sample is higher than 6,000, while a training sample with more than 6,000 posts does not reduce the estimation error significantly. Thus, 6,000 was chosen as the training sample size. In addition, around 15% of the dataset is needed to test the method performance, so the event-based dataset should contain around 8,000 posts.

The event-based dataset was constructed following the process shown in Figure 3. After estimating the size of the dataset, the authors identified the requirements regarding the events and activities considering diversity in cities, event characteristics, and their major activities.

To select the cities and activities, several criteria were identified, as listed below.

- Different cities. The authors decided to use social media posts from the two biggest cities in the Netherlands, Amsterdam and Rotterdam, as they would provide sufficient posts and cover slightly different populations as well.
- Different event characteristics. Four events were selected, as listed in Table 3.
  - A large nautical event (Sail 2015)
  - An annual national holiday (Koningsdag, or King's Day, 2016)
  - An annual celebration including a canal parade and parties (Europride 2017)
  - Football riots and championship celebration (Feyenoord 2017)
- Different activities take place during these events, including:
  - Canal parade
  - Street parties
  - Flea market
  - Fireworks
  - Riots
- Different areas in the city. The selected events include both events with activities that spread out over the whole inner city and events that are located in a smaller area.

Geo-referenced tweets and Instagram posts were then collected from selected city events according to the estimated sample size. The collection of social media posts during these events was performed through the API of social media platforms with the help of SocialGlass (http://social-glass.tudelft.nl/), an integrated system for collecting and processing social media data. The next step was to filter out spam accounts and short posts (i.e., length smaller than 30 characters) which may contain useless or insufficient information for sentiment analysis. As city events attract many foreigners, posts may contain various languages, rather than only English. To determine the sentiments of those posts, all posts were translated into English using the Google Translate API (https://cloud.google.com/translate/) which provides acceptable results compared with other translation services (*52*). The sentiment of each post was then annotated through crowd-sourcing where the sentiment of each post was determined by multiple people and the majority judgement taken as the ground truth. The crowd-sourcing operation was performed using Figure Eight (https://www.figure-eight.com/), a popular crowd-sourcing platform. Each post wass annotated using one of the terms: positive, neutral, or negative.

Table 3. City Events and Activities for Constructing Event-Based Dataset

Event	Activity	Date			Place	
		Day	Start time	End time	Area	City
Sail 2015	Sail-in parade	August 19, 2015	10:00:00	17:00:00	IJ	Amsterdam
	Sail Thank You parade	August 23, 2015	16:30:00	21:00:00	IJ	Amsterdam
	Fireworks	August 19–23, 2015	22:30:00	22:45:00	IJ	Amsterdam
King's Day 2016	Children's flea market	April 27, 2016	9:00:00	18:00:00	Vondelpark	Amsterdam
	Canal boat party	April 27, 2016	14:00:00	17:00:00	City center	Amsterdam
	King's Night	April 26–27, 2016	20:00:00	02:00:00	Rembrandtplein and Melkweg	Amsterdam
Europride 2017	Canal parade	August 5, 2017	13:00:00	17:00:00	City center	Amsterdam
	Pride park	July 28, 2017	14:00:00	23:00:00	Vondelpark	Amsterdam
	Street parties (Street Party 1)	August 4, 2017	19:00:00	02:00:00	Reguliersdwarsstraat	Amsterdam
	Street parties (Street Party 2)	August 5, 2017	16:00:00	02:00:00		
Feyenoord 2017	Lose soccer match	May 7, 2017	0:00:00	23:59:59	Feyenoord De Kuip stadium	Rotterdam
	Win the league title	May 14, 2017	0:00:00	23:59:59	Feyenoord De Kuip stadium	Rotterdam

Note: IJ is a bay area in Amsterdam.

Table 4. Common and Event-Based Datasets for Sentiment Estimation

Category	Name	Sentiment polarity	Source	# Positive	# Neutral	# Negative	# Total
Common-based	CA	Positive, Negative	University of Michigan Sentiment analysis competition	789,914	NA	788,127	1,578,041
	CB	Positive, Neutral, Negative	Niek Sanders sentiment dataset (extended)	16,146	14,004	25,379	55,537
Event-based	EA	Positive, Negative	Constructed	5,040	NA	2,029	7,069
	EB	Positive, Neutral, Negative		5,040	1,093	2,029	8,162

Note: NA = not available.

The characteristics of the common and event-based datasets used in this study are presented in Table 4. For each category there are two datasets: one with sentiment polarity of Positive and Negative, the other with the polarities Positive, Neutral, and Negative.

Experimental Setting

This section describes the set up of the experiment to test the sentiment estimation of crowds in city events using the selected methods using social media. The experiment involves multiple control variables. The following subsections first describe the values of each control variable, then introduce the experimental scenarios which combine the variable values. The final subsection describes the experiment setting applied in this experiment, that is, the training and testing sample size, the number of rounds for N-round testing.

Control Variable

The experiment is designed to test the sentiment estimation performance of selected methods when applied to city events. It therefore consists of methods and testing datasets as variables. ML-based methods also require training datasets. Thus, the training dataset acts as another variable. Moreover, as indicated in the data collection section above, two polarity sets are applied: Positive and Negative (PN), and Positive, Neutral, and Negative (PNN). Hence the sentiment polarity is also a variable for this experiment. In summary, the experiment involves control variables including: sentiment polarity, selected methods, the training data, and the testing data.

With regard to the candidate methods, two lexicon-based methods and five ML-based methods were selected, as shown in Table 2. In relation to data, both common-based and event-based datasets are used for

both training and testing. The details of these datasets have already been given in Table 4.

Scenario Design

To explore the estimation performance under different variable values, a set of scenarios was designed which combine values of those variables, as shown in Table 5 for PN polarities and Table 6 for PNN polarities.

Both Tables 5 and 6 consist of three sections (see Scenario column), investigating lexicon-based methods, ML-based methods trained using common-based data, and ML-based methods trained with event-based data, respectively. The Result column lists estimation results of each scenario, which will be discussed in the next section.

Experimental Setting

As indicated above, 1,000 samples were selected from the testing dataset for each scenario to perform 100-round testing. For ML-based methods, the training sample size was 6,000 in each round, as indicated in data collection section.

Sentiment Analysis: Findings of the Experiment

This section shows the findings and analysis of the results. They are presented and compared with and without sentiment polarity of Neutral, respectively. Within each, the discussion starts with lexicon-based methods, followed by ML-based methods. Finally, the performance of all the methods is compared.

Table 5 lists sentiment estimation results with sentiment polarity of Positive and Negative (PN). With regard to lexicon-based methods, SentiStrength reaches a similar estimation error when tested with both common-based and event-based data (mean error 0.331 and 0.322, respectively) which is better than SentiWordNet (mean error 0.407 and 0.405). Unexpectedly, ML-based methods, when trained with a common-based dataset, and tested with the event-based dataset achieved a lower minimal estimation error (mean error 0.230) than when tested with the common-based dataset (mean error 0.272). The best ML-based method appears to be LinearSVC. When ML-based methods were trained with the event-based dataset, performance tests with the event-based dataset also reach a lower minimal estimation error (LinearSVC, mean error 0.177) than when tested with a common-based dataset (NuSVC, mean error 0.453).

Likewise, Table 6 shows results for the sentiment polarity Positive, Neutral, and Negative (PNN). According to the results, the lexicon-based method SentiStrength again achieved lower estimation errors than SentiWordNet. In particular, it performed better with event-based testing dataset (mean error 0.345) than with the common testing dataset (mean error 0.451). ML-based methods showed similar patterns with the Neutral polarity (PNN) as with PN. Specifically, when trained with a common-based dataset, tests with the event-based dataset (NuSVC, mean error 0.364) performed better than when tested with a common-based dataset (LinearSVC, mean error 0.412). When trained with the event-based dataset, this pattern also holds, namely, when tested with event-based dataset (LinearSVC, mean error 0.305) the results were better than when tested with common-based dataset (SGDC, mean error 0.667.). LinearSVC reaches the lowest estimation error when both trained and tested with the event-based dataset (mean error 0.305).

When comparing all methods, ML-based methods achieved lower minimal estimation errors (when tested with event-based dataset) than lexicon-based methods. In lexicon-based methods, SentiStrength had a weaker estimation performance than SentiWordNet. For all ML-based methods, linear methods achieved more consistent results. LinearSVC reached the lowest estimation error in most scenarios, except when trained with event-based dataset and tested with common dataset, as well as vice versa, including with Neutral polarity.

When comparing sentiment estimation with and without Neutral sentiment polarity, it was found that all methods achieved lower estimation errors without Neutral polarity (PN) than with Neutral polarity (PNN).

Discussion

This section presents discussion of the results of the sentiment analysis experiment in relation to different sentiment polarities and different training and testing datasets.

With regard to different sentiment polarities, all methods show lower sentiment estimation errors when estimating the sentiments of crowds with PN, rather than using three polarities (PNN). In city events, posts sent by crowds may contain more expressions of sentiment, and stronger expressions of sentiment towards positive and negative polarities, and there may be fewer neutral posts, than in an ordinary context. This is confirmed by the distribution of sentiment in the event-based dataset constructed in this research (see data collection section, above). Therefore, estimating sentiment with PNN from these posts is more difficult, and consequently the estimation errors increase.

Following a similar reasoning, the lowest estimation error was achieved by ML-based methods trained with

Table 5. Scenario Setting of Sentiment Estimation in Polarities of Positive and Negative

		Variables											Result	
		Methods							Training data		Testing data		Error	
		Lexicon-based		ML-based					Common-based	Event-based	Common-based	Event-based		
Scenario	Name	SentiStrength	SentiWordNet	NB	SGDC	LSVC	NuSVC	SVC	CA (Pos, Neg)	EA (Pos, Neg)	CA (Pos, Neg)	EA (Pos, Neg)	Mean	Standard deviation
Lexicon methods	L-SS-PN-C	x							na	na	x		**0.331**	0.011
	L-SW-PN-C		x						na	na	x		0.407	0.013
	L-SS-PN-E	x							na	na		x	**0.322**	0.011
	L-SW-PN-E		x						na	na		x	0.405	0.011
ML methods trained with common dataset	ML-NB-PN-C-C			x					x		x		0.285	0.017
	ML-SGDC-PN-C-C				x				x		x		0.274	0.018
	ML-LSVC-PN-C-C					x			x		x		**0.272**	0.016
	ML-NuSVC-PN-C-C						x		x		x		0.459	0.069
	ML-SVC-PN-C-C							x	x		x		0.472	0.065
	ML-NB-PN-C-E			x					x			x	0.362	0.098
	ML-SGDC-PN-C-E				x				x			x	0.250	0.085
	ML-LSVC-PN-C-E					x			x			x	**0.230**	0.079
	ML-NuSVC-PN-C-E						x		x			x	0.531	0.253
	ML-SVC-PN-C-E							x	x			x	0.563	0.465
ML methods trained with event dataset	ML-NB-PN-E-C			x						x	x		0.491	0.013
	ML-SGDC-PN-E-C				x					x	x		0.490	0.013
	ML-LSVC-PN-E-C					x				x	x		0.500	0.014
	ML-NuSVC-PN-E-C						x			x	x		**0.453**	0.014
	ML-SVC-PN-E-C							x		x	x		0.505	0.015
	ML-NB-PN-E-E			x						x		x	0.184	0.011
	ML-SGDC-PN-E-E				x					x		x	0.184	0.008
	ML-LSVC-PN-E-E					x				x		x	**0.177**	0.010
	ML-NuSVC-PN-E-E						x			x		x	0.357	0.150
	ML-SVC-PN-E-E							x		x		x	0.174	0.011

Note: The Result colum denotes experiment results. Bold cells denote smallest estimation errors in the same section. The Name column denotes a combination of candidate methods, training dataset, and testing dataset, for example, L-SS-PN-C = lexicon-based method SentiStrength, in sentiment polarities of Positive and Negative (PN), tested with common-based dataset. ML-NB-PN-C-E = ML-based method NB, in sentiment polarities of Positive and Negative (PN), trained with common-based dataset, tested with event-based dataset; na = not applicable.

Table 6. Scenario Setting of Sentiment Estimation in Polarities of Positive, Neutral, and Negative

		Variables										Result	
		Methods							Training data		Testing data		Error	
		Lexicon-based		ML-based					Common-based	Event-based	Common-based	Event-based		
Scenario	Name	SentiStrength	SentiWordNet	NB	SGDC	LSVC	NuSVC	SVC	CB(Pos, Neu, Neg)	EB(Pos, Neu, Neg)	CB(Pos, Neu, Neg)	EB(Pos, Neu, Neg)	Mean	Standard deviation
Lexicon methods	L-SS-PNN-C	x							na	na	x		**0.451**	0.012
	L-SW-PNN-C		x						na	na	x		0.550	0.014
	L-SS-PNN-E	x							na	na		x	**0.345**	0.009
	L-SW-PNN-E		x						na	na		x	0.466	0.015
ML methods trained with common dataset	ML-NB-PNN-C-C			x					x		x		0.423	0.022
	ML-SGDC-PNN-C-C				x				x		x		**0.414**	0.016
	ML-LSVC-PNN-C-C					x			x		x		**0.412**	0.018
	ML-NuSVC-PNN-C-C						x		x		x		0.676	0.054
	ML-SVC-PNN-C-C							x	x		x		0.549	0.015
	ML-NB-PNN-C-E			x					x			x	0.561	0.029
	ML-SGDC-PNN-C-E				x				x			x	0.515	0.030
	ML-LSVC-PNN-C-E					x			x			x	0.526	0.026
	ML-NuSVC-PNN-C-E						x		x			x	**0.364**	0.026
	ML-SVC-PNN-C-E							x	x			x	0.643	0.003
ML methods trained with event dataset	ML-NB-PNN-E-C			x						x	x		0.678	0.012
	ML-SGDC-PNN-E-C				x					x	x		**0.667**	0.015
	ML-LSVC-PNN-E-C					x				x	x		0.671	0.015
	ML-NuSVC-PNN-E-C						x			x	x		0.710	0.022
	ML-SVC-PNN-E-C							x		x	x		0.708	0.015
	ML-NB-PNN-E-E			x						x		x	0.315	0.011
	ML-SGDC-PNN-E-E				x					x		x	0.309	0.010
	ML-LSVC-PNN-E-E					x				x		x	**0.305**	0.010
	ML-NuSVC-PNN-E-E						x			x		x	0.598	0.053
	ML-SVC-PNN-E-E							x		x		x	0.373	0.009

Note: Bold cells denote smallest estimation errors in the same section. na = not applicable.

event-based data and tested with event-based data annotated with PN sentiment polarities (LinearSVC, mean error 0.177), followed by the same method trained with common-based data and tested with event-based data with PN (i.e., LinearSVC, mean error 0.230). These observations may indicate that similarities between training and testing datasets in relation to content and context information may considerably affect the estimation performance. For instance, when training and testing data are both from events, even from different events, the texture characteristics, such as words, phrases, hashtags, emojis, and punctuation marks, may be similar, thus producing lower estimation error than training with common-based data and testing with event-based data, which are less similar.

With regard to the training dataset, with sentiment polarity of PN, the estimation error appears to be significantly distinct when ML-based methods trained with common-based or event-based data. This may also be explained by the (dis)similarity between the training and the testing dataset. The sentiments of posts in a common-based dataset are more equally distributed, while the event-based dataset contains posts with more positive or negative sentiments. Thus, the ML-based methods trained with common-based models are less biased in sentiment estimation than when trained with an event-based dataset.

Lexicon-based methods tested on both common and event-based data showed a similar error, which was worse than for the ML-based methods: lexicon-based methods take no or limited context information (e.g., weighted lexicon-based methods) into consideration, so the estimation error is increased. This is in line with the findings of Ravi and Ravi (*16*) who reviewed 161 sentiment analysis works and concluded that ML-based methods result in better accuracy than lexicon-based methods because semantic orientation provides better generality. For instance, a post such as "We are having beer on the boat! #Kingsday" is identified as a neutral post by lexicon-based methods as it is interpreted as describing a fact, but it is identified as a positive post by ML-based methods because of the context of the King's Day boat parade. This may also indicate the reason why the estimation error for lexicon-based methods tested on common and event-based data is similar; the context differences between common-based and event-based scenario data do not affect their decision.

Conclusion

City events are becoming more and more popular. Information on the sentiments of crowds is valuable when it comes to crowd management. Conventional solutions to derive such information depend on manual observations, which are expensive, prone to observation biases, and not suitable for global observations.

This paper investigates the effectiveness of methods to estimate the sentiments of crowds using social media text in the context of city events. The authors created an event-based sentiment dataset consisting of social media posts from various events and major activities. Each post was annotated with sentiment polarity of Positive, Neutral, or Negative using crowd-sourcing. This dataset has been used for the training and testing of several methods. The main objective of the research was to investigate the performance of the candidate methods using different datasets.

It was found that all candidate methods show lower estimation error with sentiment polarity of Positive and Negative, without Neutral. ML-based methods show better performance than lexicon-based methods in most situations. Specifically, the ML-based LinearSVC method reaches the minimal estimation error when trained and tested with event-based data. The findings indicate that, to predict sentiments in a crowd using social media, it is best to use ML-based method LinearSVC trained with event-based data, which achieved a mean estimation error of 0.177 approximately.

The results may be influenced by the construction bias of the event-based dataset, which are introduced by the various characteristics of selected events, and the unbalanced numbers of positive, neutral, and negative social media posts. Likewise, the bias in the common-based dataset used in this research may affect the result. Moreover, in the construction of the event-based dataset, we used the Google Translate API to translate posts in other languages into English for crowd-sourcing annotation. The accuracy of translation may introduce errors into the sentiment estimation, as posts do not follow the common way of spelling words.

In future work, the authors plan to explore more methods deriving sentiment from social media, for example, hybrid methods that integrate the lexicon-based and ML-based methods. The authors also intend to enlarge the event-based dataset by adding more diverse events and activities, and to examine the sentiment estimation performance of candidate methods across different events, or the same events in different versions. Last but not least, the sentiment estimation performance with different sentiment schemes, for example, a detailed sentiment scheme, will be investigated.

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Union Horizon 2020 Framework Programme for Research and Innovation. It is established by the Scientific Council of the ERC Grant Agreement no. 669792 (Allegro).

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: all authors; data collection: VXG; analysis and interpretation of results: VXG, WD; draft manuscript preparation: VXG, WD, AB; study supervision: SPH. All authors reviewed the results and approved the final version of the manuscript.

References

1. Still, G. K. *Crowd Dynamics*. PhD thesis. University of Warwick, U.K., 2000.
2. Zomer, L. B., W. Daamen, S. Meijer, and S. P. Hoogendoorn. Managing Crowds: The Possibilities and Limitations of Crowd Information during Urban Mass Events. In S. Geertman, J. Ferreira, Jr., R. Goodspeed, and J. Stillwell (Eds.), *Planning Support Systems and Smart Cities*, Springer, 2015, pp. 77–97.
3. Martin, A. *Factors Influencing Pedestrian Safety: A Literature Review*. PPR241, Wokingham U.K. 2006.
4. Earl, C., E. Parker, A. Tatrai, and M. Capra. Influences on Crowd Behaviour at Outdoor Music Festivals. *Environmental Health*, Vol. 4, No. 2, 2004, p. 55.
5. Poria, S., I. Chaturvedi, E. Cambria, and A. Hussain. Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis. In *2016 IEEE 16th International Conference on Data Mining (ICDM), IEEE*, 2016, pp. 439–448.
6. Poria, S., E. Cambria, N. Howard, G.-B. Huang, and A. Hussain. Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content. *Neurocomputing*, Vol. 174, 2016, pp. 50–59.
7. Cranshaw, J., R. Schwartz, J. Hong, and N. Sadeh. *The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City*. Sixth International AAAI Conference on Weblogs and Social Media, 2012.
8. Quercia, D., L. M. Aiello, R. Schifanella, and A. Davies. The Digital Life of Walkable Streets. In *Proc., 24th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee*, 2015, pp. 875–884.
9. Hasan, S., X. Zhan, and S. V. Ukkusuri. Understanding Urban Human Activity and Mobility Patterns Using Large-Scale Location-Based Data from Online Social Media. In *Proc., 2nd ACM SIGKDD International Workshop on Urban Computing, ACM*, 2013, Article No. 6.
10. Gong, V., J. Yang, W. Daamen, A. Bozzon, S. Hoogendoorn, and G.-J. Houben. *Using Social Media for Attendees Density Estimation in City-Scale Events*. IEEE Access, 2018.
11. Jiang, L., M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-Dependent Twitter Sentiment Classification. *Proc., 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, 2011, pp. 151–160.
12. Zhang, K., Y. Xie, Y. Yang, A. Sun, H. Liu, and A. Choudhary. Incorporating Conditional Random Fields and Active Learning to Improve Sentiment Identification. *Neural Networks*, Vol. 58, 2014, pp. 60–67.
13. Ortigosa, A., J. M. Martín, and R. M. Carro. Sentiment Analysis in Facebook and its Application to e-Learning. *Computers in Human Behavior*, Vol. 31, 2014, pp. 527–541.
14. Yu, Y., W. Duan, and Q. Cao. The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach. *Decision Support Systems*, Vol. 55, No. 4, 2013, pp. 919–926.
15. Pang, B., L. Lee, and S. Vaithyanathan. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. *Proc., ACL-02 Conference on Empirical Methods in Natural Language Processing, Volume 10, Association for Computational Linguistics*, 2002, pp. 79–86.
16. Ravi, K., and V. Ravi. A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications. *Knowledge-Based Systems*, Vol. 89, 2015, pp. 14–46.
17. Boiy, E., and M.-F. Moens. A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. *Information Retrieval*, Vol. 12, No. 5, 2009, pp. 526–558.
18. Li, Y.-M., and T.-Y. Li. Deriving Market Intelligence from Microblogs. *Decision Support Systems*, Vol. 55, No. 1, 2013, pp. 206–217.
19. Bravo-Marquez, F., M. Mendoza, and B. Poblete. Meta-Level Sentiment Models for Big Social Data Analysis. *Knowledge-Based Systems*, Vol. 69, 2014, pp. 86–99.
20. Thelwall, M., K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 12, 2010, pp. 2544–2558.
21. Thelwall, M., K. Buckley, and G. Paltoglou. Sentiment Strength Detection for the Social Web. *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 1, 2012, pp. 163–173.
22. Kouloumpis, E., T. Wilson, and J. D. Moore. Twitter Sentiment Analysis: The Good, the Bad and the OMG! *ICWSM*, Vol. 11, No. 538–541, 2011, p. 164.
23. Costa, H., L. H. Merschmann, F. Barth, and F. Benevenuto. Pollution, Bad-Mouthing, and Local Marketing: The Underground of Location-Based Social Networks. *Information Sciences*, Vol. 279, 2014, pp. 123–137.
24. Liu, B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, 2012, pp. 1–167.
25. Kiritchenko, S., X. Zhu, and S. M. Mohammad. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, Vol. 50, 2014, pp. 723–762.
26. Saif, H., Y. He, and H. Alani. Semantic Sentiment Analysis of Twitter. In *International Semantic Web Conference*, Springer, 2012, pp. 508–524.
27. Zhang, J.-D., and C.-Y. Chow. CoRe: Exploiting the Personalized Influence of Two-Dimensional Geographic Coordinates for Location Recommendations. *Information Sciences*, Vol. 293, 2015, pp. 163–181.
28. Aydoğan, E., and M. A. Akcayol. A Comprehensive Survey for Sentiment Analysis Tasks Using Machine Learning Techniques. In *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, IEEE, 2016, pp. 1–7.

29. Hu, M., and B. Liu. Mining and Summarizing Customer Reviews. In *Proc., 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM*, 2004, pp. 168–177.
30. Cruz, F. L., J. A. Troyano, F. Enríquez, F. J. Ortega, and C. G. Vallejo. Long Autonomy or Long Delay? The Importance of Domain in Opinion Mining. *Expert Systems with Applications*, Vol. 40, No. 8, 2013, pp. 3174–3184.
31. Blitzer, J., M. Dredze, and F. Pereira. Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification. *Proc., 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 440–447.
32. Pang, B., and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, Vol. 2, No. 1–2, 2008, pp. 1–135.
33. Martínez Cámara, E., M. Valdivia, M. Teresa, J. M. Perea Ortega, and L. A. Ureña López. Técnicas de clasificación de opiniones aplicadas a un corpus en español, Procesamiento del Lenguaje Natural, No. 47, 2011, 163–170.
34. Montoyo, A., P. Martínez-Barco, and A. Balahur. *Subjectivity and Sentiment Analysis: An Overview of the Current State of the Area and Envisaged Developments*. Decision Support Systems, Vol. 53, No. 4, 2012, pp. 675–679.
35. Thelwall, M. Heart and Soul: Sentiment Strength Detection in the Social Web with Sentistrength, 2013. *Cyberemotions: Collective Emotions in Cyberspace*. Berlin, Germany, Springer, 2013, 119–134.
36. Thelwall, M., K. Buckley, and G. Paltoglou. Sentiment in Twitter Events. *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 2, 2011, pp. 406–418.
37. Pfitzner, R., A. Garas, and F. Schweitzer. Emotional Divergence Influences Information Spreading in Twitter. *International Conference on Web and Social Media*, Vol. 12, 2012, pp. 2–5.
38. Baccianella, S., A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), Vol. 10, 2010, pp. 2200–2204.
39. Nakov, P., A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, SemEval-2016 Task 4: Sentiment Analysis in Twitter. *Proc., 10th International Workshop on Semantic Evaluation (Semeval-2016)*, 2016, pp. 1–18.
40. Gilbert, C. H. E. Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, 2014.
41. Murphy, K. P. *Naive Bayes Classifiers*. University of British Columbia, Vancouver, Vol. 18, 2006.
42. Rui, H., Y. Liu, and A. Whinston. Whose and What Chatter Matters? The Effect of Tweets on Movie Sales. *Decision Support Systems*, Vol. 55, No. 4, 2013, pp. 863–870.
43. Abdul-Mageed, M., M. Diab, and S. Kübler. SAMAR: Subjectivity and Sentiment Analysis for Arabic Social Media. *Computer Speech and Language*, Vol. 28, No. 1, 2014, pp. 20–37.
44. Tripathy, A., A. Agrawal, and S. K. Rath. Classification of Sentiment Reviews Using n-Gram Machine Learning Approach. *Expert Systems with Applications*, Vol. 57, 2016, pp. 117–126.
45. Nabil, M., M. Aly, and A. Atiya. Astd: Arabic Sentiment Tweets Dataset. *Proc., 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2515–2519.
46. Al-Azani, S., and E.-S. M. El-Alfy. Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text. *Procedia Computer Science*, Vol. 109, 2017, pp. 359–366.
47. Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts. Learning Word Vectors for Sentiment Analysis. *Proc., 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, 2011, pp. 142–150.
48. Pandey, A. C., D. S. Rajpoot, and M. Saraswat. Twitter Sentiment Analysis Using Hybrid Cuckoo Search Method. *Information Processing & Management*, Vol. 53, No. 4, 2017, pp. 764–779.
49. Lima, A. C. E., L. N. de Castro, and J. M. Corchado. A Polarity Analysis Framework for Twitter Messages. *Applied Mathematics and Computation*, Vol. 270, 2015, pp. 756–767.
50. Aston, N., T. Munson, J. Liddle, G. Hartshaw, D. Livingston, and W. Hu. Sentiment Analysis on the Social Networks Using Stream Algorithms. *Journal of Data Analysis and Information Processing*, Vol. 2, No. 02, 2014, p. 60.
51. Gurkhe, D., N. Pal, and R. Bhatia. Effective Sentiment Analysis of Social Media Datasets Using Naive Bayesian Classification. *International Journal of Computer Applications*, Vol. 99, No. 13, 2014.
52. Aiken, M., and S. Balan. An Analysis of Google Translate Accuracy. *Translation Journal*, Vol. 16, No. 2, 2011, pp. 1–3.

The Standing Committee on Traffic Flow Theory and Characteristics (AHB45) peer-reviewed this paper (19-04989).