

Itisha Gupta* and Nisheeth Joshi

Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic

<https://doi.org/10.1515/jisys-2019-0106>

Received April 24, 2019; previously published online September 20, 2019.

Abstract: This paper addresses the problem of Twitter sentiment analysis through a hybrid approach in which SentiWordNet (SWN)-based feature vector acts as input to the classification model Support Vector Machine. Our main focus is to handle lexical modifier negation during SWN score calculation for the improvement of classification performance. Thus, we present naive and novel shift approach in which negation acts as both sentiment-bearing word and modifier, and then we shift the score of words from SWN based on their contextual semantic, inferred from neighbouring words. Additionally, we augment negation accounting procedure with a few heuristics for handling the cases in which negation presence does not necessarily mean negation. Experimental results show that the contextual-based SWN feature vector obtained through shift polarity approach alone led to an improved Twitter sentiment analysis system that outperforms the traditional reverse polarity approach by 2–6%. We validate the effectiveness of our hybrid approach considering negation on benchmark Twitter corpus from SemEval-2013 Task 2 competition.

Keywords: Twitter sentiment analysis; SentiWordNet; negation handling; lexical modifier; shift approach; sentiment analysis; negation exception.

1 Introduction

User-generated opinionated data are increasing day by day through sources such as blogs, online forums, social media, and microblogging websites. Such data contain opinions about any product, topic, service, or any idea and, thus, can be effectively used for extracting valuable information from them. Among the various sources of opinionated data, Twitter is a gold mine and rich source as people tweet on each and every topic. Twitter sentiment analysis is one of the techniques for determining aggregated feelings of people from opinionated microblogging data. It plays an important role in identifying the individual's sentiment or opinion and their impact on society [6, 7, 10]. Basically, sentiment analysis techniques are broadly categorised into statistical methods, lexicon-based approach (knowledge-based methods) and hybrid approach [6]. The lexicon-based approach makes use of pre-built lexicon resources containing polarity of sentiment words such as SentiWordNet (SWN) 3.0 [4] for determining the polarity of a tweet. Lexicon-based methods are computationally efficient and scalable. However, when linguistic rules are considered, the lexicon-based approach suffers from poor recognition of sentiments. Statistical methods involve machine learning (such as Support Vector Machine, SVM) and deep learning approaches such as convolutional neural network, long short-term memory, and many more. Though both approaches require labelled training data for polarity detection, models based on deep learning have become quite popular among researchers. The hybrid approach of sentiment analysis exploits both statistical methods and knowledge-based methods for polarity detection. It inherits high accuracy from the machine learning (statistical methods) and stability from the lexicon-based approach [2]. More details regarding the different approaches of sentiment analysis are given in [6].

*Corresponding author: **Itisha Gupta**, Banasthali Vidyapith, Department of Computer Science, Tonk, India, e-mail: itishagupta07@gmail.com. <https://orcid.org/0000-0001-5270-3515>

Nisheeth Joshi: Banasthali Vidyapith, Department of Computer Science, Tonk, India. <https://orcid.org/0000-0002-9256-3825>

Most of the existing works in the literature rely on either machine learning approach [8, 12, 25, 28, 34, 41] or, recently, deep learning approach [16, 26, 31, 39, 40, 44, 48] as they proved to be more effective and accurate in sentiment classification. Such approaches have been used by researchers in various projects such as movie review classification [16, 27, 37], detection of fake news [41], detection of medicine intake from tweets [26] and many others [1, 5, 8, 18, 31, 44]. However, statistical methods require a large amount of training data and are semantically weak (predictive value for co-occurrence units is little in a statistical model). Additionally, a model trained on one domain would not perform well on another domain. Due to this, few researchers use the lexicon-based approach [2, 3, 15, 17, 19, 21, 29, 32, 35, 36, 43, 47, 49] as it does not require training data and provide consistent performance across various domains. Considering the pros and cons of both approaches, we aim to address the Twitter sentiment analysis using a hybrid approach in which we use the most frequent and general-purpose SWN lexicon resource as the first phase for generation of the feature vector and then, in the second phase, train the SVM model on SWN-based generated feature vector. To quantify the proposed hybrid approach effectiveness, we use publically available SemEval-2013 competition dataset. Details of training and testing tweets corpus are given in Section 4.

SWN, being a general-purpose lexicon contains prior polarities of words which is different from contextual polarity. Thus, the scores from SWN are context independent and might lead to misclassification, if scores are not modified accordingly. Contextual polarity can be viewed from two perspectives: one is local context and the other is global context. Local context is relatedness among words in the neighbourhood and is affected by lexical modifiers such as negation (e.g. no, not, isn't, etc.), intensifiers (e.g. very, extremely, etc.) and diminishers (e.g. barely, hardly, etc.). The global context of a word is viewed from the domain in which it occurs. For instance, the word "suck" in the movie domain expresses negative polarity, while in general it is considered a neutral word. In this work, our focus is on local contextual analysis only, especially accounting negation during score calculation through SWN. Few early works [3, 17, 24, 36] rely on SWN for sentiment classification and handle the lexical modifier negation through the most common and traditional reverse polarity approach (a word affected by negation is switched to the opposite polarity). However, this is not true in every case as negation might change the strength of polarity, too. For instance, "not excellent" is still more positive than the phrase "not good". Several studies [28, 29, 47] pointed out the inadequacy of this reverse polarity approach. For instance, Mohammad et al. [28] proved that most of the positive terms when negated tend to reverse polarity, while most of the negated terms when negated tend to change in the strength only. Motivated by the findings of [28], we aim to handle the lexical ambiguity introduced due to the presence of negation in a tweet. While working on contextual polarity, Muhammad et al. [29] in their work used the shift approach for handling the negation. Performance improvement observed by authors of [29] in their work due to the usage of shift approach inspired us to analyse the negation handling through shift approach in a deeper way. We begin with the idea of shift approach proposed by [29] and augment it with a few heuristic rules for handling the negation presence as both lexical modifier and sentiment-bearing word. Additionally, we implement rules for handling those tweets in which negation presence does not have a sense of negation (i.e. negation does not affect the polarity of the neighbourhood words). Furthermore, we update the list of negation cues (negation cues are the explicit negation words such as no, not, isn't, etc. that might affect the polarity of the neighbourhood words) used in [29] with misspelled negation words (misspelled cues are quite frequent among tweets) by looking into the Twitter word cluster created by [33]. Details of negation cases implemented as exceptions are described in Section 3. It should be noted that we use negation cues as both sentiment-bearing words and modifiers because negation cues such as "not" have real-valued polarities from SWN. Hence, based on the dominant polarity of the affected word (under negation scope), we determine which role of negation to use (sentiment-bearing or modifier). Our main contribution over the state of the art [29, 46] is handling negation through shift approach with incorporated negation exception rules in a hybrid framework of Twitter sentiment analysis.

It is important to note that the training dataset of the SemEval-2013 competition is highly unbalanced having the majority of neutral tweets. It was shown in [24] that accounting imbalance led to the performance improvement of the model. Thus, to overcome the imbalance problem of the Twitter corpus of SemEval-2013 competition, we use class weights such that minority class is given more weight and the majority class is

given less weight. It prevents the biasness of a classifier towards a more frequent class which in turn leads to performance raise. Following are the main contributions of this paper:

- We present a shift approach for handling the local contextual semantics (negation), which can capture the semantics of a word by considering neighbourhood words and then update the score of that word from SWN accordingly.
- We augment negation handling shift approach with negation exception cases, where negation presence does not necessarily mean negation.
- We show that classification performance can be improved by modelling negation during the generation of a SWN-based feature vector.
- We resolve the data imbalance problem by assigning different weights to each class according to their frequency.
- We conduct various experiments to evaluate our approach effectiveness against baselines on a benchmark Twitter dataset from SemEval-2013 task 2 competition.

The rest of the paper is organised as follows: Section 2 describes the related work done so far, Section 3 presents the framework of our hybrid approach, Section 4 presents the experimentation results, and, finally, Section 5 concludes this paper with possible future directions.

2 Related Work

Quite often, researchers addressed the problem of Twitter sentiment analysis using a supervised machine learning approach (which relies on classifiers training on features extracted from the corpus) and, recently, deep learning approach. However, such machine learning classifiers require a large amount of annotated corpora, and their performance is domain dependent. Thus, various researchers focused their study towards either lexicon-based approach, in which textual sentiment is determined through the polarities of individual words obtained from pre-built lexicon resources such as SWN [4], General Inquirer [45], MPQA (it is a subjectivity lexicon containing 8000 words labelled with their polarities and strength of polarities) [51] and Bing Liu [18], or the hybrid approach, which is a combination of both lexicon and machine learning approach. The most widely and frequently used lexicon for polarity determination is SWN 3.0, containing real value prior polarities of words. Many existing works [3, 15, 17, 21, 23, 29, 32, 35, 36] develop their model based on SWN. For instance, Jose and Choorail [21] utilised SWN with word sense disambiguation (WSD is used to determine the sense of a word according to the context in which it is used) for determining election results. They extracted tweets on Arvind Kejriwal and Kiran Bedi during the Delhi elections. They handled the negation using bootstrapping procedure [12] during the pre-processing stage and observed 1% performance improvement due to negation handling and 2.6% due to WSD. Similarly, Pamungkas and Putri [35] in their work make use of SWN for sentiment analysis of GooglePlayStore data and Twitter data. Furthermore, they presented a comparative analysis of various path-based methods of WSD such as WU-Palmer, Leacock and Chodorow (LCH), and path-based methods. Ortega et al. [32] described SSA-UO unsupervised approach presented in Semeval-2013 Task 2 for Twitter sentiment analysis. They created coarse-grained sense inventory from SWN for performing WSD and then finally performed classification using the rule-based classifier. Hogenboom et al. [17], too, utilised SWN for performing classification of English movie reviews and observed significant improvement of 6.2% in macro-averaged F1 score due to accounting negation by reverse polarity approach during SWN score calculation. They further observed an improvement of 8.0% by optimising sentiment inversion factor to -1.27 rather than using -1 .

Khan et al. [23] presented a Twitter opinion mining framework which consists of three modules including emoticon classifier, polarity classifier, and SWN classifier. A Tweet classified as neutral by the emoticon classifier is passed to the polarity classifier, which on the basis of the count of positive and negative words classifies the neutral tweets into positive, negative, and neutral. Again, the tweet classified as neutral by the polarity classifier is passed on to the final phase, i.e. SWN. They evaluated their approach on six different Twitter datasets and obtained an average accuracy of 85.7%, precision of 85.3% and recall of 82.2%.

However, they ignored the lexical ambiguity negation in their hybrid framework for Twitter sentiment analysis. Saif et al. [43] presented SENTICIRCLE lexicon-based approach which considers contextual and conceptual semantics of words while calculating their score from the lexicon. They evaluated their approach on three different Twitter datasets (including Obama McCain Debate, Health Care Reform, and Stanford Sentiment Gold Standard) using the three state-of-the-art lexicons including SWN, MPQA, and Thelwall lexicon. Results showed that their SENTICIRCLE significantly outperformed SENTISTRENGTH in accuracy, but marginally behind in F-score.

Later in 2016, another notable work was done by Muhammad et al. [29], who introduced a SMARTSA lexicon-based sentiment analysis approach which considers the local and global contextual polarity. For local context, they presented strategies for handling the negation, modifiers, and discourse structures. For global contextual polarity, they devised a domain-specific lexicon by hybridising a general-purpose SWN lexicon. They observed significant improvement in performance due to their SMARTSA lexicon-based approach. It is important to note that they presented a shift approach for negation handling unlike [43], which used reverse polarity. Furthermore, their approach, too, outperformed the state-of-the-art system SENTISTRENGTH [49]. In 2017, Asghar et al. [3] in their study presented lexicon-enhanced sentiment analysis of user reviews (drug, car, and hotel) using a rule-based framework. They integrated their rule-based framework with emoticons, modifiers, negations, and domain-specific words in order to reduce data sparsity and improvement of classification accuracy. Their work is an improvement over the work of [23] in the handling of modifiers and domain-specific words.

Most recent works [2, 15, 36] use the reverse polarity approach for negation handling in their lexicon-based sentiment analysis approach. For instance, Han et al. [15] designed SWN-based lexicon sentiment analysis framework, which uses the sentiment bias strategy in order to improve the performance of the lexicon-based sentiment analysis system. They used the small training set to learn weight and threshold parameter, which would then be used in the scoring formula of SWN lexicon. Finally, they evaluated their strategy effectiveness on the product reviews collected from Amazon. They also presented a comparison of their proposed approach with state-of-the-art SO-CAL (Semantic Orientation CALCulator) [47]. They observed that their proposed approach outperformed SO-CAL because the fixed weight used in SO-CAL is not suitable for all domains. However, for handling negation, reverse polarity strategy was used in their SWN-based lexicon sentiment analysis framework. Alaoui et al. [2] in their work presented an adaptable approach for sentiment analysis of the 2016 US election-related tweets. For this, they firstly constructed a dynamic dictionary of words with their polarities based on the several negative and positive hashtags, then classified posts into various classes, and finally, balanced the set of words before prediction. They, too, used the reverse polarity approach for handling negated context words (words that come between negation and the first punctuation mark). Pandey et al. [36] handled syntactic negation scope by a dependency parse tree and morphological negation by a prefix algorithm. Stanford Dependency Parser is used to generate the dependency parse tree, and then using the depth-first approach negation is explored. They implemented their negation handling approach in sentiment analysis of movie reviews. They observed an accuracy of 92% but used the traditional reverse polarity approach for negation accounting.

Recently, deep learning models have become quite popular and have shown considerable performance in sentiment analysis task by addressing linguistic knowledge (sentiment lexicon, negation, and intensification). For instance, a noticeable work utilising sentiment lexicons (e.g. SWN, S140, etc.) was done by Teng et al. [48] in which they computed the sentence score as a weighted sum of the prior polarity score of opinionated words. However, they used a deep learning model [bi-directional Long Short-Term Memory (LSTM)] for learning the weights and to handle semantic compositionality (intensification and negation). LSTM is a type of recurrent neural network used in deep learning field for the sequence prediction problems. They tested their model on three datasets including SemEval-2013 Twitter set, movie review (Stanford Sentiment Treebank), and mixed domains dataset. Also, their weighted sum model for handling negation and intensification is simpler than rules proposed in [47]. Qian et al. [39] proposed linguistically regularised LSTMs that, too, utilised linguistic resources such as negation, intensification (extremely, very, etc.), and lexicon to enhance the sentiment classification process. Their model addresses the shifting effect of negation and intensification by generating negation regulariser and intensification regulariser, respectively. Their proposed approach

outperformed [48] in the SST (Stanford Sentiment Treebank is a movie review dataset containing 215,154 phrases with fine grained sentiment labels) dataset. It is important to note that significant efforts are needed to develop linguistic knowledge and should be carefully incorporated into deep learning models in order to realise their potential in sentiment classification performance. On the contrary, Wang et al. [50] in their work proposed RNN-Capsule model (for each sentiment class, one capsule is created) based on Recurrent Neural Network for sentiment analysis. However, unlike [39, 48] their capsule model does not need any linguistic knowledge and can output words that can reflect dataset domain knowledge. They also evaluated their capsule approach on movie reviews and SST dataset and outperformed the state-of-the-art models [39, 48] without the incorporation of any linguistic knowledge. The above aforementioned works of deep learning used pre-trained word embedding for the vector representation of words. Word embedding is a quite popular method of creating vector representation of words in the sentiment analysis task. Ray and Chakrabarti [40] combined a deep learning approach with rule-based methods for aspect-level sentiment analysis. However, they used reverse polarity for negation handling.

Other studies make use of SWN in the hybrid approach of sentiment analysis, where the lexicon score obtained from SWN acts as input to machine learning classifiers. For instance, Kanakaraj and Guddeti [22] presented a hybrid approach by developing semantic-based feature vector through SWN with the use of WSD and then training the ensemble classifiers such as random forest, Adaboost, etc. on semantic feature vector. They observed an improvement of 3–5% with their hybrid approach over the traditional bag-of-words approach with single classifiers. However, they ignored negation handling, too. Sumanth and Inkpen [46] also described a hybrid approach for Twitter sentiment analysis of Semeval-2013 task 2 dataset. They make use of Babelfy [30] (for WSD) during polarity detection from SWN and then train the random forest classifier on semantic-based generated feature vector from SWN. While calculating the score from SWN, the local context is ignored by them. The study performed by Hung and Chen [19], too, relies on a hybrid approach by developing the hotel- and movie-based SWN lexicon from general-purpose SWN. They integrated unigrams and bigrams with vector space modelling using three classifiers [SVM, decision tree, and Naive Bayes (NB)], and results showed a performance improvement due to a WSD-based SWN lexicon. An exhaustive analysis of lexicon and machine learning approach was done by Kolchyna et al. [24]. In their study, they described various techniques for implementing lexicon and machine learning approaches. Furthermore, they tried the various combinations of lexicons and found an improvement in performance when the lexicon is enhanced with emoticons and slangs. They also implemented a hybrid approach (ensemble of lexicon and machine learning approach) on a benchmark dataset (SemEval-2013 Task 2) which produced a more accurate classification. At last, they observed an improvement of 7% due to the use of cost-sensitive SVM classifier for handling class imbalance problem. In their study, negation is handled by the reverse polarity approach. Recently, Carrillo-de-Albornoz et al. [8] performed feature level sentiment analysis in e-health-related forums. They did an exhaustive analysis of various features such as semantic, lexical, syntactic, and word embedding. They used the vector-based models (reduce data sparseness) for the representation of words and presented a comparative analysis with the bag-of-words model (does not capture similarities among words) in combination with other features. They evaluated combinations of different features through various state-of-the-art machine learning classifiers such as NB, Sequential Minimal Optimization, random forest, and Vote (ensemble of different classifiers). They observed the best results with word embedding in combination with lexical and sentiment-based features. Furthermore, they suggested the importance of objective information in a health-related domain. However, they implemented negation in the form of binary features (presence or absence of negation) only and observed that their negation features did not help in polarity prediction. Thus, they stated that a sophisticated mechanism is needed for negation accounting rather than just using binary negation features.

Most of the above aforementioned works [19, 22, 23, 32, 35, 46] use a plain bag-of-words without considering relatedness among words during score calculation from SWN. However a few works [2, 3, 15, 17, 24, 36, 43] consider negation with score calculation from the lexicon, but again they use traditional reverse polarity approach for negation handling which has been proved inadequate by [28, 29] in their work. Thus, we can conclude that most of the existing works in Twitter sentiment analysis focused on handling the global context (depend on the domain in which word appears) of a word either through WSD [21, 22, 35, 46] or by creating a domain-specific lexicon [19, 29, 32]. The above mentioned works either ignore the impact of local

context (relatedness among words in the neighbourhood) or handled the local context in an improper way [2, 3, 8, 17, 24, 36, 38, 43, 49]. The first computational model for negation handling was presented by [38], but again they used the reverse polarity for negation accounting.

Few works such as Taboada et al. [47] used a fixed amount of 4 to shift towards opposite polarity for handling negation in their lexicon-based approach rather than a complete reversal. However, their shift approach for negation did not match with human judgment. Muhammad et al. [29], too, presented a shift approach for negation handling, but unlike [47] their shifting is not by a fixed amount. They shift score of affected words according to the score of negation cues based on the fact that negation cues themselves have a score in SWN. Since we aim to use SWN in our hybrid approach; we begin with shift approach proposed in [29] and improve it by augmenting it with a few heuristics to handle the cases where negation presence does not necessarily mean negation. Also, we aim to model the negation handling shift strategy in a hybrid framework of Twitter sentiment analysis because it is a challenging task to improve the classification performance using a hybrid approach, considering the impact of contextual valence shifters. Hence, in this work, we focus on performing a three-way classification of microblog data (tweets) using a hybrid technique, by considering the impact of local contextual modifiers especially negation.

Our main focus is to handle the lexical modifier negation and show how much negation accounting through shift approach with incorporated negation exception rules helps in improvement of classifier performance in the hybrid framework.

3 Methodology

In this section, we will describe our proposed hybrid approach for Twitter sentiment analysis considering contextual valence shifter negation. The proposed framework (overall workflow is shown below in Figure 1) is implemented in various phases and starts with publically available SemEval-2013 dataset (Section 4.1 presents the corpus details). Tokenisation and POS (part of speech such as noun, verb, adverb, and many more) tagging are done using twitter-specific CMU POS tagger (it is a Twitter specific tagger which captures the Twitter specific entities such as URLs, punctuations, usernames, etc. as separate entities) [13]. Also, tweet normalisation (cleaning and removal of noise from a tweet) is conducted for all the tweets in the dataset. Finally, SWN is used in conjunction with negation handling procedure to generate feature vectors which would act as input to the classification process.

3.1 Data Pre-Processing

First of all, we use CMU POS Tagger for performing tokenisation and part-of-speech tagging of our input text (Section 4.1 presents the corpus details). The result of the CMU tagger is a bag-of-words with their POS tags. The reason for using CMU tagger and not any other tagger is that this tagger is specifically created for tweets considering their unstructured nature. For instance, consider a tweet “Looks like Andy the Android may have had a little too much fun yesterday. <http://t.co/7ZDEfzEC>”. The output of CMU tagger would be “Looks like Andy the Android may have had a little too much fun yesterday. <http://t.co/7ZDEfzEC> V P ^ D ^ V V V D A R A N N, U”.

Thus, it is clearly visible from the above example that CMU tagger captures the Twitter-specific entities such as URLs as a separate entity and tags it by POS tag “U”. The next step is to perform tweet normalisation, i.e. cleaning and normalising of unstructured noisy tweets such as removal of usernames, URLs, stop words, punctuations, misspell replacement, or slang replacement. Such tweet normalisation operations make them ready for SWN lexical resource because SWN is a formal lexicon and does not contain the misspelled, slang, and elongated words. Thus, for performing such tweet normalisation, we use a few modules from our previous work [14] in which we have proposed a tweet normalisation system that would clean and normalise the tweets.

The last and most important part of pre-processing is negation scope detection, which is to mark all the words which come under the negation scope so that we can easily identify words in negated context during shifting of scores from SWN. Negation scope detection is itself another challenging task. Thus, we explore the

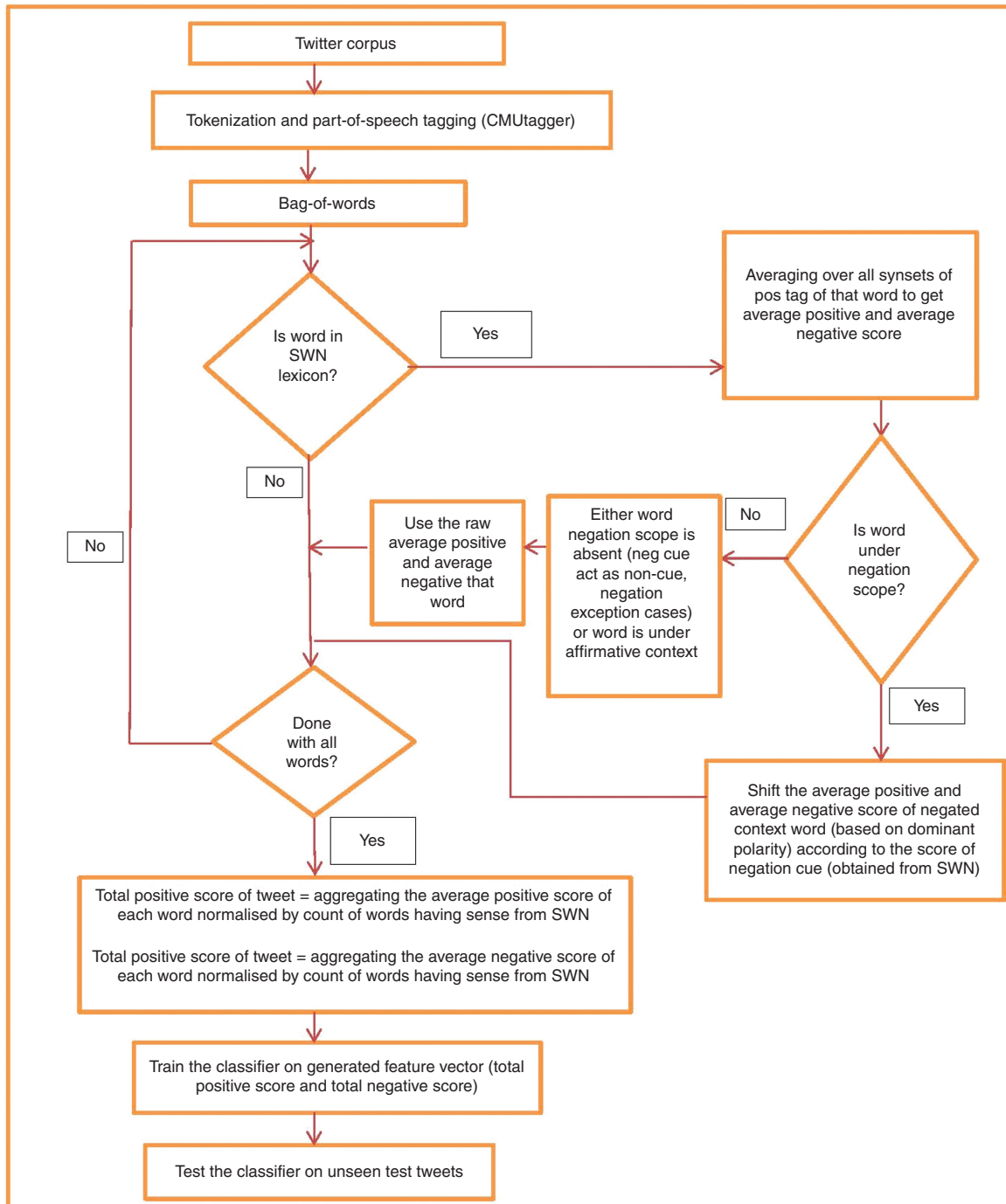


Figure 1: The Overall Workflow of our Proposed Hybrid Framework.

possible ways of scope detection in the existing literature. Most of the early works [2, 3, 21, 24, 28] use simple and traditional policy in which they mark all the words between negation and the first punctuation marks. However, such an approach would mark all words under scope between negation and first punctuation mark irrespective of whether it is polar or not. We keep our negation scope detection naive and simple because complex approaches for scope detection such as machine learning, deep learning [9], complex rules, and compositional semantics parsing [20] are either computationally intensive or require annotated corpora. Our negation scope detection idea is to use a static window of five words but to consider the impact of linguistic constructs such as conjunctions (but) and punctuations which might expire the negation scope before

window termination. Furthermore, we only consider adjectives, verbs, nouns, and adverbs during scope marking.

Moreover, after analysing the tweets having negation cues, we came across various tweets in which negation cue is present but there is no negation sense. Often, a non-cue occurs as a determiner (POS tag is “D”) or exclamation (“No! I am not ready”). Such tweets are having negation either in the form of phrases such as no one, not only, by no means, etc. or in the form of negative rhetoric questions (for instance, “I’m in Petrolia. The sun’s out, isn’t that bright enough for you?”). We call such tweets as negation exception cases, and in our work, we augment shift approach for negation handling with a few heuristic rules as existing researches lack on providing an appropriate way of handling negation exception cases described below.

- Exception case 1: When negation is a part of an expression (that does not carry negation sense) such as not only, not just, no one, at no times, no wonder, no question, not to mention, and by no means. For example, “I am not sure about #DeMonetisation but in such a rhetoric there is **no one** as good as him”. In this tweet negation cue “no” is present but it is a part of the phrase “no one”, and there is no sense of negation. Thus, we consider that negation scope is absent, and no negation handling would be done in that case.
- Exception case 2: When negation word happens in negative rhetoric questions. For example, “**Isn’t it** already a failure when people still talk about the problems of #DeMonetisation and not its successes after 50 days! Think”. Generally, the rhetoric question is identified by two conditions: one is the presence of a question mark, and other is the presence of negation in the first three words of text [11, 20]. However, in the case of an unstructured tweet, it is not possible to capture the negation in the rhetoric form with those conditions. Thus, we manually analyse the POS tags of negation cues and their neighbouring words in rhetoric questions. Based on those POS tags, we finally come up with the patterns as shown in Table 1, where negation cue does not necessarily mean negation. Table 1 presents the negation exception cases based on POS tags of negation cue and the next two tokens. For instance, if the POS tag of a negation cue in a tweet is “V” (verb) and the next two tokens tags are “D” (determiner) and “A” (adjective), respectively, then no negation handling would be done (no score shifting). However, if the POS tag of a negation cue is either “D” (determiner) or “!”, then we will not check the next two tokens tags. For example, consider a tweet phrase “No way it’ll be better than Taylor Allderdice but I’m still pumped for it. @RealWizKhalifa”. Here negation word “no” whose POS tag is “D” is not affecting the polarity of the opinionated word “better”, i.e. “no” acts as non-cue. Thus, negation handling would not be done in this tweet.

3.2 SWN-Based Feature Vector Generation

This is the main phase of our hybrid approach where we make use of SWN lexicon with negation accounting to create a feature vector for classifier training. We implement this phase in two sub-modules, namely, getting the raw score and shifting the raw score (if the word is under negated context).

3.2.1 Getting the Raw Score of Word from SWN

In this sub-module, firstly we retrieve the score of each token of a pre-processed tweet (obtained from the first phase) from SWN by averaging over all the synsets of the target token, normalised by the length of synsets (as shown in Figure 1). We ignore the objective score here because we aim to show the negation impact on positive and negative polarity. Thus, we got two scores of each available word in SWN: one is an average positive

Table 1: Negation Exception Patterns from Negative Rhetoric Questions.

Negation cue (t)	Next token (t + 1)	Next token (t + 2)
D	—	—
!	—	—
V	D	A
V	O	—

score, and other is an average negative score. Note that we also obtain the score of negation cue from SWN considering the negation cue as both modifier and sentiment-bearing word.

It is worth noting that negation cues in apostrophe form such as isn't, can't, didn't, etc. have no scores in SWN. Only the cue "not" with POS tag adverb ("r") has a score in SWN. Thus, we use the score of "not" for apostrophe cues ending with "n't" considering the fact that, at the end the expansion of negation apostrophe cues ending with "n't" would be "not". For instance, the expansion of cue "isn't" would be "is not".

3.2.2 Shifting the Raw Score, If Word under Negated Context

This sub-module is the central component of our hybrid approach. Here, we are going to shift the average positive and negative score (obtained through the previous submodule) of a target word (under negated context) with respect to the score of negation cues from SWN. Note that no shifting (use the raw scores) will be done for the words under affirmative context (as shown in Figure 1). Following are the rules for shifting approach:

- If the dominant polarity of the target word is positive (dominant polarity is positive if the average positive score is greater than the average negative score), then do not use the average positive score of the target word and shift the negative score of the target word by an amount equal to the negation cue negative score (example 3, Table 2). The negation cue positive score would be the updated positive score of the target word. Let x and y be the positive and negative scores of the negation cue, respectively, and a and b be the average positive and negative score, respectively, for the target word. If $a > b$, the dominant polarity is positive, and the negative score " b " would be updated to " $b + y$ ". Now x is the new positive score of the target word, and $b + y$ is the new negative score of the target word.
- If the dominant polarity of the target word is negative (dominant polarity is negative if the average positive score is less than the average negative score), then we do not use the negation cue score in the aggregation process. Hence, the updated negative score of the target word is "0.0", and the positive score will remain the same (example 2, Table 2). The reason for not using negation cues score in aggregation is that we get undesirable results (example 1, Table 2). As seen in example 1, the phrase "not bad" would still remain negative (-0.608) if we use negation cue score.
- If the average positive score of the target word is equal to the average negative score, then no shifting is done.

Thus, each tweet is represented as a feature vector of size 2:

- The total positive score of a tweet, obtained by aggregating the positive score of each word (determined after applying shifting model of negation coupled with SWN), normalised by the count of the number of words having the sense from SWN.
- The total negative score of a tweet, obtained by aggregating the negative score of each word (determined after applying shifting model of negation coupled with SWN), normalised by the count of the number of words having the sense from SWN.

Table 2: Negation Adjustment Examples.

Score	Negation cue	Target word	Final score of target word after shifting
Example 1: (incorrect shifting)	Not	Bad	
Avg. positive score	0.0	0.017	0.017
Avg. negative score	0.625	0.660	0.625 (0.017-0.625 = -0.608)
Example 2: (correct shifting)	Not	Bad	
Avg. positive score	0.0	0.017	0.017
Avg. negative score	0.625	0.660	0.0 (0.017-0.0 = 0.017)
Example 3: (correct shifting)	Won't	Respect	
Avg. positive score	0.0	0.437	0.0
Avg. negative score	0.625	0.0	0.625 + 0.0 = 0.625

3.3 Classification and Testing

This is the last phase of our hybrid approach in which SWN-based generated feature vector acts as input to the supervised learning classifier. In this phase, we train the SVM classifier on the generated feature vector from the previous phase and make predictions on the unknown test set. We chose the SVM classifier because it is proven previously to be robust and used by the top-performing NRC (National Research Council)-Canada team [28] of SemEval-2013 task 2. We used the RBF (Radial Basis Function) kernel SVM (Section 4.2 presents details about parameter optimisation of SVM) since our feature vector is of size 2 only and linear kernel SVM [28] proved to be effective on a larger feature space. A detailed description of various experimentations done using SVM is given in Section 4.2.

4 Experiments and Results

In this section, we describe the details of our corpus and experimentations done in order to evaluate the effectiveness of our hybrid approach to twitter sentiment analysis. We implemented our framework in python3.4 and used the Scikit-learn for the SVM implementation.

4.1 Corpus

We used the benchmark Twitter corpus from SemEval-2013 competition Task 2: Sentiment Analysis in Twitter. This task includes two subtasks: (a) term-level classification and (b) message-level classification. We are focused on the subtask b that is classifying a message into three classes: positive, negative, and neutral (message-level polarity classification). The organisers of this SemEval-2013 conference provided training data, development data, and testing data, whose details are described below in Table 3.

4.2 Experiments

To begin with experiments, firstly we optimise the C and gamma parameters of SVM classifier using 10-fold cross validation and grid search. We obtained the optimised value of C to be 10 and gamma to be 0.1. We used macro-averaged recall [42] score as primary evaluation metrics because in case of class imbalance, it is more robust than standard F1 and accuracy. The macro-averaged recall is computed by averaging over the recall of positive, negative, and neutral classes. The other secondary evaluation metrics that we use are accuracy, macro-averaged precision, and F1-score. We conducted two types of experiments. One is to present the comparison of our hybrid approach with baseline. Second is to assess the contribution of local context handling (negation) through shift approach.

Table 4 presents the results of the baseline and our hybrid approach using SVM, evaluated using 10-fold cross validation on SemEval-2013 training data. Here, the baseline is the most frequent classifier which

Table 3: SemEval-2013 Task 2 Data Statistics.

Dataset	Positive	Negative	Neutral	Total
Training	2978 (37.14%)	1162 (14.49%)	3878 (48.37%)	8018
Development	483 (34.97%)	280 (20.28%)	618 (44.75%)	1381
Test	1306 (40.84%)	484 (15.13%)	1408 (44.03%)	3198

Table 4: Ten-fold Cross-Validation Result on Training Data.

System	Accuracy	Average recall	Average F1
Baseline	48.37	33.33	21.73
Our system	58.67	57.65	55.39

always predicts the most frequent class. In our training set, the neutral class is the most frequent; hence, the baseline always predicts the neutral class.

We also reported the evaluation metrics such as accuracy, precision, recall, and F1 for all the three classes. It is worth noting that optimised parameters of SVM were obtained through grid search on the training data only, but the SVM is trained on both training and development sets and then tested on the unseen tweets test set. Table 5 describes the complete result obtained by our proposed hybrid approach on SemEval-2013 test set.

We conducted the second set of experiments in order to determine the significance of adding local contextual semantics into a plain bag-of-words model by accounting negation through shift approach. To observe the impact of handling local contextual semantics, we run SVM classification algorithm again on the test set but with varying SWN-based feature vector (as shown in Figure 2) that is plain bag-of-words (disregarding negation), bag-of-words enhanced through reverse polarity of negation handling, and, finally, bag-of-words enhanced using our shift approach (with augmented negation exception rules).

Table 6 presents results of experiments conducted on test set for different negation processing strategies. It is clear from the result that our shift approach to negation handling led to an improvement of 6 percentage points in macro-averaged recall and F1-score. Also, there is an improvement of 2 percentage points in macro-averaged precision and 3.7 percentage points in accuracy over the reverse polarity. Observe that reverse polarity is still better than not accounting for negation. Furthermore, negation exception rules offer an additional gain of 1 percentage point in macro-averaged F1-score and precision. This slight performance

Table 5: Proposed Hybrid Approach Result on Test Data for Each Class.

System	Precision	Recall	F1-score
Positive	69	52	59
Negative	42	50	46
Neutral	61	70	65

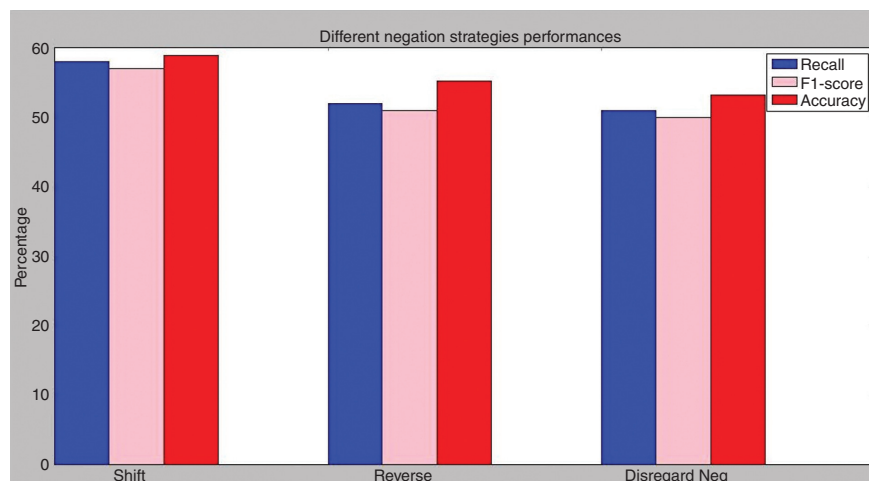


Figure 2: Comparison of Different Negation Strategies.

Table 6: The Result on SemEval-2013 Test Set for Different Negation Processing Strategies.

System	Macro-avg. recall	Macro-avg. F1	Macro-avg. precision	Accuracy
Disregarding negation	51	52	50	53.2
Reverse polarity	52	51	55	55.2
Shift polarity without negation exception rules	58	56	56	58.3
Shift polarity with negation exception rules	58	57	57	59.8

Table 7: The Macro-Averaged F1-Score on SemEval-2013 Test Set for Positive and Negative Classes Only.

System	Tweet test set (macro-avg. F1-score)
Baseline 1 (majority classifier)	28.9%
Baseline 2 (first sense)	47%
NRC-Canada (unigram features only) [28]	39.61%
Babelify [46]	50.75%
Our system	52.5%

improvement is due to the fact that there are a very low number of instances (1–2%) in SemEval-2013 corpus which have negation but there is no sense of negation.

Our main aim is to show how much SWN-based features (obtained from SWN by negation accounting through shift approach) help in Twitter sentiment analysis without the help of any other features. Thus, Table 7 presents a comparison of our SWN-based contextual features with only unigram features generated by NRC-Canada team [28] (top performing team in message-level subtask of SemEval-2013). Such unigram features include POS-based features, elongated words, capital words, emoticon features, negation features, unigram Twitter-specific emoticon, and hashtag lexicon features. We observe an improvement of 12.89 percentage points in macro-averaged F1-score on the SemEval-2013 test set over the NRC-Canada unigram only features. This demonstrates the contribution of negation handling alone in performance improvement because we are not using any of the unigram features of NRC-Canada. Table 7 also presents a comparison of our approach with state-of-the-art [46] that, too, used a hybrid approach on SemEval-2013 corpus. They used Babelify for WSD during SWN-based feature vector creation but did not handle negation. Though we use averaging of all synsets, rather than word sense disambiguation algorithm [46], we are able to outperform the result [46] by 1.75 percentage points. Also, Table 7 shows the results of baseline1 (majority classifier always predicts the most frequent class which is positive in this case) and baseline 2, when the first sense is considered from SWN during score calculation. We present results for this fold of experiments in terms of macro-averaged F1-score only because it was the primary evaluation metric for SemEval-2013 task 2. Though the classification was performed for all the three classes, macro-averaged F1 score was calculated only for positive and negative classes as shown in Table 7.

Our intuition is not to show our approach as best performing in SemEval-2013 subtask B (message-level polarity). We only intend to show the contribution of negation accounting through shift approach in getting considerable performance.

Moreover, we also show the performance of the proposed system (shown below in Table 8) using additional features such as POS features (the number of occurrences of each unique POS tags), n-grams (the presence or absence of unigrams and bigrams), Twitter-specific (such as number of elongated words, number of words with all letters capitalised, number of exclamation marks, question marks, presence or absence of emoticons, and many more) and cluster features. The CMU Twitter NLP tool provides 1000 clusters (generated with Brown clustering algorithm on 56 million English tweets) which are an alternate representation of tweets because Twitter data are highly unstructured in nature and contain many misspellings, slangs, and unusual expressions. Thus, we represent the presence or absence of each token from each of the 1000 clusters and number of occurrences of each cluster as features. The importance of cluster features is to perform generalisations through the counting of cluster occurrences that include similar type words. Our proposed system obtained considerable performance with additional features, too.

Table 8: Proposed Hybrid Approach Performance with Additional Features on Test Data for Each Class.

System	Precision	Recall	F1-score
Positive	84	59	69
Negative	55	58	57
Neutral	67	83	74

5 Conclusion and Future Work

In this work, we have presented hybrid approach (a combination of lexicon and machine learning approach) to Twitter sentiment analysis which focuses on the contribution of negation handling in performance raised through shifting approach during SWN-based feature calculation. We have also implemented the cases as exceptions where negation cue is present but there is no negation sense in order to avoid misclassification. We observed substantial performance improvement of 6 percentage points in macro-averaged recall and F1 due to shift approach augmented with negation exception rules over the traditional and most common reverse polarity approach of handling negation on the publically available benchmark SemEval-2013 Twitter dataset. We did not use any other features such as n-grams, punctuations, emoticons, and POS-based and other lexicon-based features during the comparison of our proposed system with the state-of-the-art [46] and NRC-Canada unigram only features [28] so as to show that improvement in classification performance is due to accounting negation alone, not with the aid of any other features. In the future, we would assess the impact of other local contextual shifters such as intensifiers and diminishers. We would also evaluate the impact of discourse-based information such as conjunctions and conditionals on the classification performance.

Bibliography

- [1] M. S. Akhtar, A. Ekbal, S. Narayan, V. Singh and E. Cambria, No, that never happened!! Investigating rumors on Twitter, *IEEE Intell. Syst.* **33** (2018), 8–15.
- [2] I. E. I. Alaoui, Y. Gahi, R. Messoussi and Y. Chaabi, A novel adaptable approach for sentiment analysis on big social data, *J. Big Data* **5** (2018), 12.
- [3] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim and I. A. Khan, Lexicon-enhanced sentiment analysis framework using rule-based classification scheme, *PLoS One* **12** (2017), e0171649.
- [4] S. Baccianella, A. Esuli and F. Sebastiani, SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, *LREC* **10** (2010), 2200–2204.
- [5] L. Branz and P. Brockmann, Sentiment analysis of Twitter data: towards filtering, analysing interpreting social networking data, in: *Proceedings of the 12th ACM International Conference on Distributed Event-Based Systems*, pp. 238–241, Hamilton, New Zealand, June, 2018.
- [6] E. Cambria, Affecting computing and sentiment analysis, *IEEE Intell. Syst.* **31** (2016), 102–107.
- [7] E. Cambria, S. Poria, A. Gelbukh and M. Thelwall, Sentiment analysis is a big suitcase, *IEEE Intell. Syst.* **32** (2017), 74–80.
- [8] J. Carillo-de-Albornoz, J. R. Vidal and L. Plaza, Feature engineering for sentiment analysis in e-health forums, *PLoS One* **13** (2018).
- [9] L. Chen, Attention-based deep learning system for negation and assertion detection in clinical notes, *Int. J. Artif. Intell. Appl.* **10** (2019), 1–9.
- [10] M. Ebrahimi, A. Z. Yazdavar and A. Sheth, Challenges of sentiment analysis for dynamic events, *IEEE Intell. Syst.* **32** (2017), 70–75.
- [11] U. Farooq, H. Mansoor, A. Nongaillard, Y. Ouzrout and M. A. Qadir, Negation handling in sentiment analysis at sentence level, *JCP* **12** (2017), 470–478.
- [12] Y. Garg and N. Chatterjee, Sentiment analysis of Twitter feeds, in: *International Conference on Big Data Analytics*, pp. 33–52, New Delhi, India, December, 2014.
- [13] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan and N. A. Smith, Part of speech tagging for Twitter: annotation, features, and experiments, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 42–47, 2010.
- [14] I. Gupta and N. Joshi, Tweet normalization: a knowledge-based approach, in: *2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS)*, Dubai, UAE, December, 2017.
- [15] H. Han, Y. Zhang, J. Zhang, J. Yang and X. Zou, Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias, *PLoS One* **13** (2018), e0202523.
- [16] A. Hassan and A. Mahmood, Deep learning approach for sentiment analysis of short texts, in: *2017 3rd International Conference on control, automation and robotics (ICCAR)*, pp. 705–710, Nagoya, Japan, April, 2017.
- [17] A. Hogenboom, P. Van Iterson, B. Heerschop, F. Frasincar and U. Kaymak, Determining negation scope and strength in sentiment analysis, in: *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2589–2594, Anchorage, AK, USA, 2011.
- [18] M. Hu and B. Liu, Mining and summarizing customer reviews, in: *Proceedings of the 10th ACM SIGKDD International conference on Knowledge discovery and Data Mining*, pp. 168–177, Seattle, WA, USA, August, 2004.

- [19] C. Hung and S. J. Chen, Word sense disambiguation based sentiment lexicons for sentiment classification, *Knowl.-Based Syst.* **110** (2016), 224–232.
- [20] L. Jia, C. Yu and W. Meng, The effect of negation on sentiment analysis and retrieval effectiveness, in: *Proceedings of 18th ACM Conference on Information and Knowledge Management*, pp. 1827–1830, Hong-Kong, China, November, 2009.
- [21] R. Jose and V. S. Choorailil, Prediction of election result by enhanced sentiment analysis on Twitter data using word sense disambiguation, in: *2015 International Conference on Control Communication and Computing India (ICCC)*, Trivandrum, India, 2015.
- [22] M. Kanakaraj and R. M. R. Guddeti, NLP based sentiment analysis on Twitter data using ensemble classifiers, in: *2015 3rd International Conference on Signal Processing, Communication and Networking*, pp. 1–5, Chennai, India, March, 2015.
- [23] F. H. Khan, S. Bashir and U. Qamar, TOM: Twitter opinion mining framework using hybrid classification scheme, *Decis. Support Syst.* **57** (2014), 245–257.
- [24] O. Kolchyna, T. T. P. Souza, P. Treleaven and T. Aste, Twitter sentiment analysis: lexicon method, machine learning method and their combination, in: *Handbook of Sentiment analysis in Finance*, 2016.
- [25] E. Kouloumpis, T. Wilson and J. Moore, Twitter sentiment analysis: the good, the bad and the OMG!, in: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, July, 2011.
- [26] D. Mahata, J. Friedrichs, R. R. Shah, J. Jiang and E. Cambria, Detecting personal intake of medicine from Twitter, *IEEE Intell. Syst.* **33** (2018), 87–95.
- [27] Q. I. Mahmud, A. Mohaimen and M. S. Islam, A Support Vector Machine mixed with statistical reasoning approach to predict movie success by analysing public sentiments, in: *2017 20th International Conference of Computer and Information Technology*, pp. 1–6, Dhaka, Bangladesh, December, 2017.
- [28] S. M. Mohammad, S. Kiritchenko and X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets, in: *Second Joint, Conference on Lexical and Computational Semantics (SEM)*, pp. 321–327, Atlanta, USA, June, 2013.
- [29] A. Muhammad, N. Wiratunga and R. Lothian, Contextual sentiment analysis for social media genres, *Knowl.-Based Syst.* **108** (2016), 92–101.
- [30] R. Navigli and S. P. Ponzetto, BabelNet: building a very large multilingual semantic network, in: *Proceedings of the 48th Annual Meetings of the ACL*, pp. 216–225, Uppsala, Sweden, 2010.
- [31] D. Nguyen, K. Vo, D. Pham, M. Nguyen and T. Quan, A deep architecture for sentiment analysis of news article, in: *International Conference on Computer Science, Applied Mathematics and Applications*, pp. 129–140, Berlin, Germany, June, 2017.
- [32] R. Ortega, A. Fonseca and A. Montoyo, SSA-UO: Unsupervised Twitter sentiment analysis, in: *Second Joint conference on Lexical and Computational Semantics (*SEM)*, pp. 501–507, Atlanta, Georgia, USA, 2013.
- [33] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider and N. A. Smith, Improved part-of-speech tagging for online conversational text with word clusters, in: *Proceedings of NAACL: HLT*, pp. 380–390, Atlanta, Georgia, June, 2013.
- [34] A. Pak and P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, *LREC*, pp. 1320–1326, Valletta, Malta, May, 2010.
- [35] E. W. Pamungkas and D. G. P. Putri, Word sense disambiguation for lexicon-based sentiment analysis, in: *9th International Conference on Machine Learning and Computing*, Singapore, 2017.
- [36] S. Pandey, S. Sagnika and B. S. P. Mishra, A technique to handle negation in sentiment analysis on movie reviews, in: *2018 International Conference on Communication and Signal Processing (ICCSP)*, pp. 737–743, Chennai, India, April, 2018.
- [37] B. Pang, L. Lee and S. Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, in: *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, pp. 79–86, Philadelphia, July, 2002.
- [38] L. Polanyi and A. Zaenen, Contextual valence shifters, computing attitude and affect in text: *Theor. Appl.* **20** (2004), 1–9.
- [39] Q. Qian, M. Huang, J. Lei and X. Zhu, Linguistically regularized LSTMs for sentiment classification, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1679–1689, Vancouver, Canada, July, 2017.
- [40] P. Ray and A. Chakrabarti, A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis, *Appl. Comput. Inform.* (2019). doi: 10.1016/j.aci.2019.02.002.
- [41] J. C. S. Reis, A. Correia, F. Murai and A. Veloso, Supervised learning for fake news detection, *IEEE Intell. Syst.* **34** (2019), 76–81.
- [42] S. Rosenthal, N. Farra and P. Nakov, SemEval-2017 task 4: sentiment analysis in Twitter, in: *Proceedings of 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 502–5018, Vancouver, Canada, August, 2017.
- [43] H. Saif, M. Fernandez, Y. He and H. Alani, Senticircle for contextual and conceptual semantic sentiment analysis of twitter, *European Semantic Web Conference*, pp. 83–98, Springer, Cham, 2014.
- [44] W. Souma, I. Vodenska and H. Aoyama, Enhanced news sentiment analysis using deep learning method, *J. Comput. Soc. Sci.* **2** (2019), 33–46.
- [45] P. J. Stone, D. C. Dunphy and M. S. Smith, The general Inquirer: a computer approach to content analysis, *MIT Press*, 1996.
- [46] C. Sumanth and D. Inkpen, How much does word sense disambiguation help in sentiment analysis of micropost data? in: *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media analysis*, pp. 115–121, Lisbon, Portugal, 2015.

- [47] M. Taboada, J. Brooke, M. Tofigoski, K. Voll and M. Stede, Lexicon based methods for sentiment analysis, *Comput. Linguist.* **37** (2011), 267–307.
- [48] Z. Teng, D. T. Vo and Y. Zhang, Context-sensitive lexicon features for neural sentiment analysis, in: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1629–1638, Austin, Texas, November, 2016.
- [49] M. Thelwall, K. Buckley and G. Paltoglou, Sentiment strength detection for the social web, *J. Am. Soc. Inf. Sci. Technol.* **63** (2012), 163–173.
- [50] Y. Wang, A. Sun, J. Han, Y. Liu and X. Zhu, Sentiment analysis by capsules, in: *Proceedings of the 2018 World Wide Web Conference*, pp. 1165–1174, Lyon, France, April, 2018.
- [51] T. Wilson, J. Weibe and P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in NLP*, pp. 347–354, British Columbia, Canada, 2005.