



## SEMANTIC GRAPH BASED TERM EXPANSION FOR SENTENCE-LEVEL SENTIMENT ANALYSIS

Mohammed Maree <sup>1)</sup>, Mujahed Eleyat <sup>2)</sup>

<sup>1)</sup> Department of Multimedia Technology,  
Faculty of Engineering and Information Technology,  
Arab American University, Palestine,  
Email: mohammed.maree@aaup.edu, <https://www.aaup.edu/mohammed.maree>

<sup>2)</sup> Department of Computer Systems Engineering,  
Faculty of Engineering and Information Technology,  
Arab American University, Palestine  
Email: mujahed.eleyat@aaup.edu

### Paper history:

Received 28 May 2020  
Received in revised form 10 September 2020  
Accepted 17 October 2020  
Available online 30 December 2020

### Keywords:

Semantic graph;  
Sentiment analysis;  
POS-based Term expansion;  
Machine learning;  
Term Prior Polarity.

**Abstract:** The semantic orientation (also referred to as prior polarity) of a word plays an important role in automatic sentence-level sentiment analysis. Several approaches have been proposed wherein a lexicon of words marked with their polarities is exploited to infer the meaning of sentences. However, relying on prior word polarity may produce inaccurate decisions. This is because we may find negative-sentence sentiments that include words with positive prior polarities or vice versa. In this article, we propose an approach to sentence-level sentiment analysis that exploits knowledge encoded in heavy-weight semantic graphs to assist in discovering the meaning of a word in the context of the sentence where it appears. In this context, we build contextual semantic networks for indexing sentences and expand them with semantically/lexically-relevant terms in an attempt to disambiguate the meanings of word mentions in sentences. In order to verify the effectiveness of the proposed approach, we have developed a prototype system using a real-world dataset that contains 46830 sentiment sentences along with a gold-standard that comprises 10000 movie reviews that are labelled under five sentiment categories (very negative, negative, neutral, positive, very positive). Findings indicate that enriching the semantic graphs of sentiment sentences with NOUN-based synonyms and hypernyms has improved the overall quality of baseline sentiment analysis techniques.

*Copyright © Research Institute for Intelligent Computer Systems, 2020.  
All rights reserved.*

## 1. INTRODUCTION

Sentiment Analysis (a.k.a. Opinion Mining) can be defined as the process of automatically identifying, analyzing and understanding people's opinions, appraisals, and emotions toward various products, services, and topics [1, 2]. It indeed plays a crucial role in understanding the perception (user-generated reviews, comments, tweets, and so on) of users about products or services that they use or deal with. Automatically, understanding such content saves time, cost and effort required by organizations to manually review and analyze users' impressions about what they serve. Over the past few years, greater attention has been given to developing

techniques that attempt to automate the process of analyzing users' generated content and accurately identify its meaning. The goal in this context is to provide analytic tools that can assist organizations better understand their users, and improve the quality of their services accordingly [3, 4]. However, among the major drawbacks of conventional sentiment analysis approaches is the treatment of words in sentences in an isolated-context manner, ignoring the word polarity on the one hand and its context-sensitiveness on the other. To address this issue, we propose a new approach to sentence-level sentiment analysis that exploits knowledge captured in large-scale semantic graphs to assist in

discovering the meaning of a word in the context of the sentence where it appears. Consequently, and unlike conventional approaches, we construct subject-predicate-object triplet-based contextual semantic networks to represent each sentence, and expand it with semantically and lexically relevant terms in an attempt to disambiguate the meanings of word mentions. As we further demonstrate in the experiments section, we tested the proposed technique against a number of sentiment analysis baseline techniques using two real-world datasets that comprised: 1) a total of 46830 sentiment sentences obtained from Twitter, and 2) a reference a gold-standard that contains 10000 movie reviews which are labelled under five sentiment categories (very negative, negative, neutral, positive, very positive). In particular, we experimentally explored the impact of expanding words in sentiment reviews using a variety of semantically/lexically related expansion terms, including hypernyms, hyponyms, synonyms and meronyms. It is important to mention here that our proposed expansion approach assigns different weights to words in sentences to identify representative sets of candidate expansion words. To do this, we employ a word re-weighting technique to evaluate the level of informativeness of each word in a sentence and select those that appear to significantly contribute to the meaning of the sentence. Accordingly, we summarize the main contributions of this research work as follows:

- Constructing subject-predicate-object triplet-based contextual semantic graphs to represent sentiment sentences. Our aim in this context is to build semantic networks that encode the semantic and taxonomic relations that link words in sentiment sentences, and enable expanding them with semantically-relevant terms that are obtained from large-scale ontology. We experimentally demonstrate that expanding words in sentiment sentences with different types of expansion candidates (such as synonyms, hypernyms, etc.) have an impact on the overall quality of the algorithms used for identifying the sentiment orientation of sentiment sentences.
- Utilizing a word weighting technique that selects representative words (among a given sentiment sentence) that appear to significantly contribute to the meaning of the sentence and expand them with multiple semantically/lexically related words. We demonstrate that the incorporation of the part of speech (POS) tagger plays a crucial role in the word expansion process as various word POS categories result in producing different accuracy levels. In particular, we experimentally demonstrate that the expansion

of words with their corresponding synonyms that belong to NOUN POS category has produced higher accuracy sentiment prediction results when compared to other POS categories.

The rest of this paper is organized as follows. In section 2, we review a number of sentiment analysis approaches and highlight their strengths and weaknesses. Section 3 provides the formal definitions of the terms and techniques used in our research work, and also presents our methodology for sentiment analysis. Section 4 introduces the experiments that we have conducted to evaluate the quality of the proposed techniques. In this section, we also provide details about the datasets used and the various algorithms that we employed to measure the effectiveness of the discussed sentiment analysis techniques. Finally, in section 5, we draw the conclusions and detailed steps for future works.

## 2. RELATED WORK

With the ever increasing growth in online users' generated content, it has become more important for companies to identify and discover people's opinions about the products and services they offer to their users [5, 6]. As reported in [3], conventional approaches have relied on company-related textual documents that are normally gathered over a 10-years' time period in an attempt to assist organizations better analyze their current and future risks. However, most of these approaches are either based on the bag-of-words model, or utilize word-level embedding techniques, which are imprecise due to the fact that these models treat words in a context-isolated manner, ignoring the semantic as well as contextual aspects of words and their mentions in sentences [7]. To address these issues, newer sentiment analysis approaches have been proposed [8-12]. The goal in this context is to analyze and understand the huge amount of online textual information that users generate in the form of product or service reviews. Such information can be broadly categorized into facts and opinions, where facts represent objective expressions such as entities and their properties, while opinions express subjective views that describe people's sentiments, appraisals or feeling [13]. The use of sentiment analysis techniques has proved to play a crucial role in this context [2, 14-16]. As stated by Sharma and Dey [17], with sentiment analysis, classifying textual documents under various sentiment categories, such as positive or negative, has become possible. In the same research direction, Jeong et. al. have emphasized the fact that commercial firms need to pay particular attention to customer voices to provide them with new or improved products. The authors reported that approaches to research,

development and marketing have placed considerable emphasis on customer needs analysis. The goal in this context is to co-create value with customers, since firms are directly or indirectly engaged with customers who want to satisfy their product and service needs [11]. To practically realize sentiment analysis techniques in real-world application domains and scenarios, various natural language processing (NLP), text analysis and mining techniques are employed. The aim of these techniques is to preprocess user-generated content, segment the content into document, sentence, word/term, or aspect levels, and analyze them to produce sentiment scores, and extract opinions. Among the solutions that have been proposed to sentiment analysis is the work proposed in [18]. In their research work, Bessalov et. al. proposed to use high order n-grams for the purpose of classifying the sentiment orientation of a given text at article level, under the assumption that longer phrases are likely to be less ambiguous in terms of their polarity. However, unfortunately, as stated by the authors, this proposal comes at a very high computational cost as it requires to model tri or a greater number of n-grams leading to an extremely large parameter space associated to n-grams. Other works reported in [19-21] employed supervised learning approaches to sentiment classification of Twitter tweets. These approaches have relied on the exploitation of training data for sentiment classifier learning, such as Naïve Bayes, Maximum Entropy, and Support Vector Machines. The used training data was obtained either using emoticons associated with users' tweets, or by collecting consensus from the results returned by the sentiment detection websites.

However, these approaches suffer from two major drawbacks. First, since user tweets can be classified under multiple domains of interest, training data is always dependent on the domain of interest, and thus for each domain, new training data will be required. Second, generalizing the proposed approaches on other application domains is difficult due to its high dependence on subjective data that is obtained using sources that are not always available in other application domains, such as those that offer the possibility of adding emoticon to each sentiment sentence. In a similar line of research, the authors of [22-24] have proposed utilizing word features, such as word n-grams [22, 23] and semantic patterns [24] to predict the sentiment orientation of user-generated reviews. However, as we have discussed earlier in section 1, these approaches ignore the contextual semantic relations among word mentions in sentiment sentences. In other words, they treat words in sentences and other derived semantic aspects in an isolated-context manner, leading to deviate the semantic as well as sentiment orientation of

sentiment sentences. In the same line of research, Jeong et. al. [11] proposed an approach to analyze user-generated content on social medial platforms. The authors proposed a model that comprised three main components. These are Latent Dirichlet Allocation (LDA) based topic modeling, sentiment analyzer, and an algorithm called opportunity algorithm. Using the proposed model, the topic about each product from customers' perspective is identified using LDA. Then, the degrees of importance of each topic were then calculated using topic satisfaction scores and sentiment matrix. Based on the obtained scores, the model produces the product improvement opportunity value and direction using customer-centered views. However, despite the contributions made by this model, there is a number of weaknesses that still hinder the full exploitation of the model in practical settings as reported by the authors. Among these weaknesses is the imprecise identification of the standard degrees of importance and satisfaction. In addition, the proposed approach was applied on only one example target product, limiting the reproducibility as well as accuracy of the model when applied on other new application domains.

Other researchers have proposed to employ deep learning techniques for analyzing the sentiment orientation of a given set of product reviews. Among the works in this domain is the model proposed by Zhao et al. [25]. In their proposed framework, called Weakly-supervised Deep Embedding (WDE), the authors employed review ratings as a training resource for a sentiment classifier. They used Convolutional Neural Network (CNN) to develop their WDE-CNN model, and Long Short-Term Memory (LSTM) for developing the WDE-LSTM to extract feature vectors from review sentences. As reported by the authors, the proposed model was experimentally evaluated using Amazon dataset that was mainly focused on three product domains: digital cameras, cell phones, and laptops. The prediction accuracy obtained using the WDE-CNN model was 87.7%, and using the WDE-LSTM model was 87.9%. These results reflect the quality of the proposed model and demonstrate deep learning models can produce high quality sentiment analysis results, especially when compared with the bag of words and other similar baseline techniques. Xiong et al. [26] proposed another model called Multi-level Sentiment-enriched Word Embedding (MSWE), which employs a Multi-layer perceptron neural network to model word-level sentiment orientation and CNN to model tweet level sentiment information. The authors tested the proposed model using SemEval2013 and Context-Sensitive Twitter datasets, which are the benchmark datasets for sentiment classification task. As reported by the

authors, experimental results demonstrated high quality sentiment analysis results on both datasets. However, it is important to point out that the main drawback of current models that employ deep learning techniques lies in the fact that they are less efficient than other sentiment analysis approaches that do not require preparing a training dataset to train the model. Requiring a training dataset in this context means that for every new product review domain a new training dataset will be needed. In addition, contextual relations among sentence terms are ignored in these model, and as such degrading the quality of the prediction process.

To address these issues, we propose to model sentiment sentences using subject-predicate-object triplet-based contextual semantic networks that encode the semantic and taxonomic relations that link words in sentiment sentences. This model also enables expanding words in sentences with semantically-relevant terms that are obtained from extrinsic large-scale ontologies and knowledge resources that provide a rich source of semantic information about generic as well as domain specific terms that appear in sentiment sentences. In addition, we propose to employ a word weighting technique that selects representative words that appear to make contribution to the meaning of the sentence and expands them with multiple semantically/lexically related words. In this context, we substantiate our argument by experimentally demonstrating that some words in a given sentiment sentence have little contribution to meaning of sentences and accordingly can be either retained without expansion or even removed from the sentence. Example of such words are stopwords and other supportive words that have low term frequencies across all sentiment sentences in the dataset. We also argue that various word POS categories can result in deviating the accuracy of sentiment analysis techniques. Therefore, we have experimentally explored the impact of the incorporation of expansion words that belong to various POS categories including *NOUN* and *ADJECTIVE*. We further detail this step in the Section 4, where we demonstrate that *NOUN*-based enrichment produced higher accuracy sentiment prediction results.

### 3. FORMAL DEFINITIONS AND PROPOSED METHODOLOGY

In this section, we first start by formally defining the terms “*sentiment analysis*”, “*word weighting*”, and “*contextual semantic network*”. We then proceed with introducing our proposed method for sentiment analysis and highlight the main steps involved in the proposed approach.

### 3.1 FORMAL DEFINITIONS

As we have discussed in the introduction section, sentiment analysis is the process of analyzing user-generated content that is normally expressed in the form of reviews, tweets, comments, etc. to infer users’ perceptions on the products and services that they use. Formally, we define this process as follows:

Let  $D = \{d_i \mid i \in [1 - M]\}$  be a set of documents, where each  $d_i$  is composed of a sequence  $S = \{s_1, s_2, \dots, s_n\}$  of sentiment sentences or reviews, and each  $s_i$  consists of a set of words  $W = \{w_j \mid j \in [1 - M]\}$ . Formally, we define the process of sentiment analysis as follows:

**Definition 1: Sentiment Analysis:** Given a set of documents  $D$  with a sequence of sentences  $S$ , the sentiment analysis algorithm first identifies the set of head concepts  $H = \{h_j \mid i \in [1 - M]\}$  among  $w_j \in W$  that will be further processed to determine the sentiment orientation of each  $s_j \in S$ . The algorithm then identifies the semantic orientation of each  $h_j \in H$  to predict and assign the sentiment orientation of each  $s_j \in S$ . It is important to mention here that, unlike conventional approaches such as the work proposed in Hu & Liu [27], we do not limit the head concepts identification procedure to adjective words only as this may lead to deviating the semantic direction of sentences from the actual norm a sentiment sentence depicts. Therefore, we construct contextual semantic networks using the concepts in  $H$  in an attempt to disambiguate the semantic orientation of each  $h_j \in H$ . In our context, head concepts are identified using the word weighting algorithm that is formally described below:

**Definition 2: Word Weighting:** given a set of words  $W = \{w_j \mid j \in [1 - M]\}$ , the algorithm filters  $w_j \in W$  in the following manner:

- First, all  $w_j$  words that belong to the set of stopwords  $\{st_1, st_2, \dots, st_N\}$  that are defined in  $ST = \{st_n\}_{n=1}^N$  are removed from  $W$ .
- Second, the algorithm uses the filtered set  $W$  to identify the set of head concepts  $H$ . To do this, remaining words in  $W$  are assigned to their part of speech (POS) categories as follows. We start with AJ, NN, ADV, and VV words. It is important to mention that although we are using multiple forms of a word, we prioritize them starting with nouns. Our decision was substantiated by the fact that noun-based expansion candidates proved to produce high accuracy sentiment prediction results compared to other words obtained using other POS categories. Therefore, despite the fact that adjectives are used to express users’ perceptions as reported in Kirubaharan et al, [28], we argue that expanding words in sentiments sentences with noun-based synonyms and hypernyms leads to higher accuracy results. This step is carried out using the Stanford CoreNLP, where each  $w_j$  is classified under the grammatical category that

it belongs to. Namely, the POS-Tagger module is used to accomplish this task. Then, we exploit knowledge encoded in WordNet ontology to find semantically and lexically related terms to each word in  $H$ . Specifically, we use the below weighting formulate:

$$W(h_i) = \begin{cases} 1 + \log_{10} tf(h_i) \times Syn(h_i) \times Hyp(h_i) & \text{if } tf_{h_i, s_i} > 0 \\ 0, & \text{othersize} \end{cases}, \quad (1)$$

where,

- $tf(h_i)$ : is the total number of occurrences of each  $h_i$ .
- $Syn(h_i)$ : is the number of synonymous terms of each  $h_i$ . We may find zero, one or more synonyms for each  $h_i$ . For instance, the words “first-class, fantabulous” are synonymous of the term “excellent”.
- $Hyp(h_i)$ : is the number of hypernyms of each  $h_i$ . By moving one level up in the hierarchical structure of the ontology, we may find none, one or more hypernyms for each  $h_i$ . For instance, we don’t find any hypernyms for the term “excellent”, while we find that both terms “advantage” and “vantage” are synonymous parents of the term “good”.
- Third, the algorithm expands the initial set of head concepts  $H$  with those obtained using step 2 above. In this context, an expanded set of head concepts  $H_{exp}$  is returned. This set is then used as input for the contextual semantic networks construction algorithm.

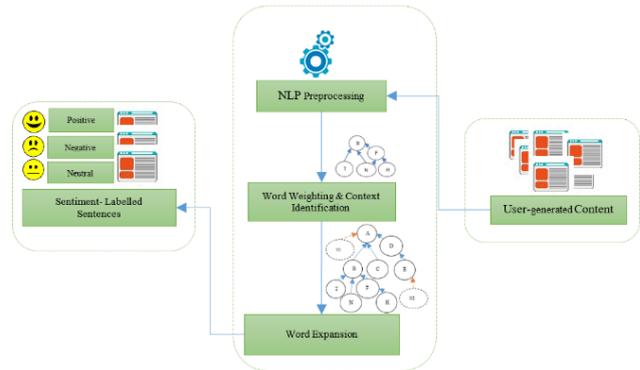
**Definition 3: Contextual Semantic Network:**

Given a set of head concepts  $H_{exp}$ , the algorithm constructs a contextual semantic network through linking  $h_i$  and  $h_j$  with both semantic and taxonomic relations  $R = \{r_j | j \in [1 - N]\}$  that are derived from the knowledge graphs of the exploited ontology. Formally, we define a contextual semantic network as a labeled graph  $G = (N, R, L_e)$ , where  $N$  is a set whose elements are called nodes from  $H_{exp}$ , and  $R$  is a set of unordered pairs  $\{r_i, r_j\}$  of relations that are used to represent the edges of the graph, and  $L_r$  an edge labeling function  $L_r: R \rightarrow L_r$ . As we mentioned earlier in this section, before adding a new node to  $N$ , it is important that the algorithm identifies the correct sense of each  $h_i$  in order to disambiguate it from other senses that may be defined for the same concept in the exploited ontology. To do this, we utilize a semantic path matching technique that finds the similarity between the path that is constructed from the words of each sense of  $h_i$  and those that are defined in each  $s_j \in S$ . A semantic path in this context is defined as a directed path between nodes  $h_i$  and  $h_j$  with a sequence  $h_0 r_1 h_1 r_2 h_2 \dots r_{n-1} h_{n-1} r_n h_n$  ( $n > 0$ ) where  $h_0 r_1 h_1, h_1 r_2 h_2, \dots, h_{n-2} r_{n-1} h_{n-1}, h_{n-1} r_n h_n$  are triples in  $G$ . The length of the path is  $n$ .

In the next section, we introduce the main steps that are involved in our proposed method for sentiment analysis. We provide an overall overview of the main components of the proposed model and further discuss their details.

**3.2 PROPOSED METHODOLOGY AND SYSTEM OVERVIEW**

As depicted in Figure 1, to conduct the sentiment analysis task, we carry out a number of steps that are organized in the following sequence.



**Figure 1 –Sentiment Analysis Steps Involved in the Proposed Approach**

First, for a given dataset of user-generated reviews, we utilize a sequence of NLP steps, including stopwords removal, depluralization, and POS tagging. These steps are important at this phase as they filter out the words in each sentiment sentence and retain refined sets of words from where we can identify and select head concepts. Next, we apply the word weighting and context identification algorithm to select head concepts among other words in each sentiment sentence. After identifying head concepts, the algorithm identifies the context of each concept and enriches it with additional semantically and taxonomically related terms, including synonyms and hypernyms of each term as formally discussed in Section 3.1. Then, the contextual semantic networks construction starts to build semantic networks for each sentiment sentence in the dataset. Relations and enrichment candidate concepts are acquired based on the knowledge graphs of the exploited ontology. Finally, using the constructed contextual semantic networks, the semantic orientation of each sentence is determined in order to assign its sentiment orientation. We use two groups of sentiment labels to represent sentiment orientations of sentences. These are 1) a three-class labels group with positive, neutral, and negative labels, and 2) a five-class labels group with very positive, positive, neutral, negative and very negative labels. Each labels group is used with the appropriate dataset as we demonstrate later in the experimental section.

## 4. EXPERIMENTAL SETUP AND EVALUATION RESULTS

In this section, we present the experimental steps that we have carried out in order to evaluate the quality of the proposed sentiment analysis techniques. Before we proceed to presenting the details of each of these steps, we would like to point out that we have implemented the proposed solutions using Java programming language that is installed on a 64-bits Windows 10 O.S with 16 GB of memory. In addition, we have used other software packages and libraries, including Stanford CoreNLP<sup>1</sup>, RapidMiner<sup>2</sup> and LightSide<sup>3</sup>. The goal of using these packages is to assist in gathering dataset elements and applying some pre-processing, filtration and storage tasks as we will demonstrate later in this section.

### 4.1 DATASET ACQUISITION AND PREPARATION

Acquiring a dataset that comprises hundreds or even thousands of user-generated sentiment sentences is not an easy task. However, with the availability of recent software tools such as RapidMiner which offers developers an authorized access to Twitter tweets via an API key, this task has become easier and more efficient. We have used this software as one of our sources for gathering sentiment reviews using the following queries:

1. **Query No. 1:** “Facebook”. We have used this query keyword to collect 10000 tweets about Facebook. The language of the tweets is “en” and the results type is “recent or popular”.
2. **Query No. 2:** “Virtual Reality”. We have used this bi-gram query to collect 10000 tweets about virtual reality. The language of the tweets is “en” and the results type is “recent or popular”.
3. **Query No. 3:** “Karate”. We have used this uni-gram query to collect 10000 tweets about users’ perception of Karate as one of the well-known martial arts. The language of the tweets is “en” and the results type is “recent or popular”.
4. **Query No. 4:** “Data Mining”. We have used this bi-gram query and collected 3106 tweets about Data Mining domain. The language of the tweets is “en” and the results type is “recent or popular”.
5. **Query No. 5:** “Kid Games”. We have used this bi-gram query to collect 10000 tweets about kid games. The language of the tweets

is “en” and the results type is “recent or popular”.

6. **Query No. 6:** “Jackie Chan”. We have used this bi-gram query to collect 3724 tweets about this famous celebrity. The language of the tweets is “en” and the results type is “recent or popular”.

Using the above-mentioned queries, we gathered a dataset that comprises 46830 tweets about six different topics that fall under different domains of interest. In addition to this manually acquired dataset, we have also used a publicly-available dataset that consists of 10000 sentiment reviews about movies that has been used as part of LightSide’s dataset. We have used these datasets for evaluation purposes as we demonstrate in the next section.

### 4.2 EXPERIMENTAL EVALUATION

First, we start with the movie reviews dataset and use it as our gold-standard for evaluating the quality of state-of-the-art sentence-level sentiment analysis techniques. In particular, we use the CoreNLP-RAW and Aylien sentiment analyzers, and compare their quality with WordNet-based sentiment analyzer. We would like to point out that we have developed two versions of the WordNet-based analyzer. These are: 1) *NOUN*-based Synonyms-Hypernyms-enriched WordNet analyzer (henceforth referred to as NSHW) and 2) *ADJECTIVE*-based Synonyms-Hypernyms-enriched WordNet analyzer (henceforth referred to as ASHW). Our aim in this context is to explore the impact of enriching sentiment sentences with different POS-based synonyms and hypernyms. In order to run our experiments, we have employed various training models to precisely predict sentiment sentences in the gold-standard dataset. Table 1 shows the various training models that we have utilized along with their accuracy results.

**Table 1. Accuracy results for the utilized training models**

Model	Metric	Value
Naïve Bayes	Accuracy	<b>0.7745</b>
	Kappa	<b>0.5491</b>
Logistic Regression	Accuracy	0.7587
	Kappa	0.5174
Support Vector Machine	Accuracy	0.73157
	Kappa	0.4631

<sup>1</sup> <https://stanfordnlp.github.io/CoreNLP/>

<sup>2</sup> <https://rapidminer.com/>

<sup>3</sup> <http://ankara.lti.cs.cmu.edu/side/>

In Table 2, we present the confusion matrix results for each of the utilized models. As we can see in this table, the confusion matrix for the Naïve Bayes model demonstrates a higher level of correctly identified negative as well as positive sentiment sentences.

**Table 2. Confusion Matrix for the utilized training models**

Model	Act\Pred	neg	pos
Naïve Bayes	neg	<b>4212</b>	1119
	pos	1285	<b>4046</b>
Logistic Regression	neg	4075	1256
	pos	1317	4014
Support Vector Machine	neg	3944	1387
	pos	1475	3856

Based on the results shown in both Tables 1 and 2, we have decided to employ Naïve Bayes model as our training model for predicting the sentiment orientation of the sentiment sentences in our dataset. In other words, since we have empirically found the Naïve Bayes model to be more accurate in predicting classes of sentiment sentences, we use it in the next experiments to predict the sentiment classes of sentiment sentences in our dataset.

As we have discussed earlier in this section, we compare the accuracy of the predictions produced by the employed sentiment analysis techniques as shown in Table 3.

**Table 3. Comparison between the results produced by the employed sentiment analysis techniques**

Dataset (Sentiment Sentences per Query)	Model Prediction Accuracy			
	CoreNLP	Aylien	NSHW	ASHW
Data Mining	83.88%	82.01%	92.82%	91.72%
Jackie Chan	88.61%	76.02%	89.84%	88.35%
Facebook	83.25%	48.83%	88.30%	88.01%
Karate	77.61%	61.00%	85.05%	85.15%
Kid Games	78.81%	67.16%	88.79%	81.23%
Virtual Reality	85.35%	52.49%	91.06%	90.98%

As we can see in Table 3, for almost all sentiment analysis results, the NSHW algorithm has proved to outperform the other algorithms (CoreNLP-RAW, Aylien and ASHW) by producing the highest sentiment analysis and prediction scores. This indeed applies to all datasets, except for the dataset that comprised sentiment sentences about the query Karate, where the algorithm produced a slightly

lower level of accuracy when compared to the ASWH algorithm. However, we can see that the overall quality of the sentiment analysis can be improved by exploiting the proposed model (in both versions NSHW and ASHW). Therefore, we can conclude that the semantic modeling and representation of sentiment sentences in the form of contextual semantic networks, and expanding them with additional enrichment candidates can play an important role in improving the accuracy of baseline sentiment prediction techniques. However, we would like to point out that our current approach is still hindered by the semantic knowledge incompleteness limitation that is inherent in WordNet ontology. Therefore, despite the precision of this ontology (as it was manually constructed by domain experts), it still lacks a lot of concepts as well as semantic and taxonomic relations. We plan to overcome this problem by further exploiting additional large-scale ontology such as YAGO, which comprised millions of entities that are linked using various types of lexical and semantic relations.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, a new approach to sentiment analysis has been proposed. Firstly, we have reviewed a number of existing sentiment analysis techniques and highlighted their main strengths and weaknesses. Then, we introduced the details of our proposed approach, and formally defined the sentiment analysis problem, in addition to describing the main steps that are involved for carrying out the sentiment analysis task. We have also detailed the various components of the proposed approach and highlighted the features that characterize each component. In order to evaluate the quality of our proposed model, we have used a dataset that comprised 46830 tweets about six different topics that fall under different domains of interest. We have used a second gold-standard dataset as a reference for evaluating the precision of the newly assigned sentiment labels to each sentence using our proposed model. Findings indicate that modeling sentiment sentences in the form of contextual semantic networks and enriching them with semantically and taxonomically related terms can play a significant role in improving the quality of baseline sentiment analysis and prediction techniques. We have also demonstrated that the enrichment with candidate concepts that belong to various POS categories can result in producing different accuracy results. In particular, we found that NOUN-based expansion has proved to outperform other POS-based expansion methods. However, we have also found that our current proposed approach is still hindered by the fact that the used ontology is limited in terms

of its domain coverage as well as accuracy in identifying the semantic polarity and contextual paths of words in sentiment sentences. Therefore, we plan to carry out further experiments using an additional semantic resource and knowledge graphs, such as YAGO ontology which comprises millions of entities that are linked using various types of semantic and taxonomic relations.

## 6. REFERENCES

- [1] S. Bhatia, M. Sharma, K. K. Bhatia, and P. Das, "Opinion target extraction with sentiment analysis," *International Journal of Computing*, vol. 17, no. 3, pp. 136-142, 2018.
- [2] V. Kharde and S. S. Sonawane, "Sentiment analysis of twitter data: a survey of techniques," *International Journal of Computer Applications*, vol. 139, no. 11, pp. 5-15, 2016.
- [3] C.-H. Du, M.-F. Tsai, and C.-J. Wang, "Beyond word-level to sentence-level sentiment analysis for financial reports," *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1562-1566: IEEE.
- [4] H. J. VL and S. Duraisamy, "Novel pre-processing framework to improve classification accuracy in opinion mining," *International Journal of Computing*, vol. 17, no. 4, pp. 234-242, 2018.
- [5] T. Cu, H. Schneider, and J. V. Scotter, "How does sentiment content of product reviews make diffusion different?," *Journal of Computer Information Systems*, vol. 59, no. 2, pp. 127-135, 2019.
- [6] M. Poturak, S. Softic, "Influence of social media content on consumer purchase intention: Mediation effect of brand equity," *Eurasian Journal of Business and Economics*, vol. 12, no. 23, pp. 17-43, 2019.
- [7] B. Agarwal, N. Mittal, P. Bansal, S. Garg, "Sentiment analysis using common-sense and context information," vol. 2015, Article ID 715730, p. 30, 2015. <https://doi.org/10.1155/2015/715730>
- [8] K. Schouten, F. Frasincar, and F. de Jong, "Ontology-enhanced aspect-based sentiment analysis," *Proceedings of the International Conference on Web Engineering*, 2017, pp. 302-320.
- [9] N. Zainuddin, A. Selamat, and R. J. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Applied Intelligence*, vol. 48, no. 5, pp. 1218-1232, 2018.
- [10] M. Horakova, "Sentiment analysis tool using machine learning," *Global Journal on Technology*, vol. 2015, no. 5, pp. 195-204, 2015.
- [11] B. Jeong, J. Yoon, and J.-M. J. I. J. o. I. M. Lee, "Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis," *International Journal of Information Management*, vol. 48, pp. 280-290, 2019.
- [12] N. K. Singh, D. S. Tomar, and A. K. Sangaiah, "Sentiment analysis: a review and comparative analysis over social media," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 97-117, 2020.
- [13] B. Liu, "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, vol. 2, no. 2010, pp. 627-666, 2010.
- [14] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008. <https://doi.org/10.1561/15000000011>
- [15] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," *Mining Text Data*: Springer, 2012, pp. 415-463. [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)
- [16] B. S. Rintyarna, R. Sarno, C. Faticah, "Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context," *International Journal of Information and Decision Sciences*, vol. 12, no. 1, pp. 75-101, 2020. <https://doi.org/10.1504/IJIDS.2020.104992>
- [17] A. Sharma and S. Dey, "A document-level sentiment analysis approach using artificial neural network and sentiment lexicons," *ACM SIGAPP Applied Computing Review*, vol. 12, no. 4, pp. 67-75, 2012.
- [18] D. Beshpalov, B. Bai, Y. Qi, and A. Shokoufandeh, "Sentiment classification based on supervised latent n-gram analysis," *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 2011, pp. 375-382.
- [19] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg!," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011, pp. 538-541.
- [20] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," *Proceedings of the International Semantic Web Conference*, 2012, pp. 508-524.
- [21] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," *Proceedings of the 23rd International*

*Conference on Computational Linguistics: Posters*, 2010, pp. 36-44.

- [22] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," *Proceedings of the International Conference on Discovery Science*, 2010, pp. 1-15.
- [23] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2010, [http://www.lrec-conf.org/proceedings/lrec2010/pdf/385\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf).
- [24] H. Saif, Y. He, M. Fernandez, and H. Alani, "Semantic patterns for sentiment analysis of Twitter," *Proceedings of the International Semantic Web Conference*, 2014, pp. 324-340.
- [25] W. Zhao *et al.*, "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Transactions on Knowledge Data Engineering*, vol. 30, no. 1, pp. 185-197, 2017.
- [26] S. Xiong, H. Lv, W. Zhao, and D. Ji, "Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings," *Neurocomputing*, vol. 275, pp. 2459-2466, 2018.
- [27] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 168-177.
- [28] V. Gnana Kirubakaran, P. Elumalai, and A. Yovan Felix, "Online product ranking based on reviews," *ARPN Journal of Engineering and Applied Sciences*, vol. 12, no. 13, pp. 4097-4100, 2006.



**Mohammed Maree** received the Ph.D. degree in Information Technology from Monash University. He began his career as a Research and Development Manager with gSoft Technology Solution, Inc. He was the Director of research and QA with Dimensions Consulting Company. Subsequently, he joined the Faculty of Engineering and Information Technology (EIT), Arab American University, Palestine (AAUP), as a full-time Lecturer. From September 2014 to August 2016, he was the Head of the Multimedia Technology Department, and from September 2016 to August 2018, he was the Head of the Information Technology Department. In addition to his work at AAUP, he worked as a Consultant for SocialDice and Dimensions Consulting companies. He is currently the Head of the Multimedia Technology Department, Faculty of Engineering and Information Technology, AAUP. He has published articles in various high-impact journals and conferences, such as ICTAI, Knowledge-Based Systems, IEEE Access and the Journal of Information Science. He is currently a Committee Member and a Reviewer of several conferences and journals. He has supervised a number of master's students in the fields of knowledge engineering, data analysis, information retrieval, natural language processing, and hybrid intelligent systems.



**Mujahed Eleyat** is an Assistant Professor of Computer Science at the Arab American University in Palestine. Moreover, he has been the Dean of the Faculty of Engineering and Information Technology since August 2019. He received his B.A. in Electrical Engineering from Birzeit University and his Master and Ph.D degrees in Computer Science from University of Arkansas and Norwegian University of Science and Technology respectively. Dr. Eleyat areas of expertise include high performance computing, parallel programming and embedded systems.