Research Article

Sentiment Analysis in Spanish for Improvement of Products and Services: A Deep Learning Approach

Mario Andrés Paredes-Valverde,[1] Ricardo Colomo-Palacios,[2] María del Pilar Salas-Zárate,[1] and Rafael Valencia-García[1]

[1]*Departamento de Informática y Sistemas, Universidad de Murcia, 30100 Murcia, Spain*
[2]*Computer Science Department, Østfold University College, Holden, Norway*

Correspondence should be addressed to María del Pilar Salas-Zárate; mariapilar.salas@um.es

Sentiment analysis is an important area that allows knowing public opinion of the users about several aspects. This information helps organizations to know customer satisfaction. Social networks such as Twitter are important information channels because information in real time can be obtained and processed from them. In this sense, we propose a deep-learning-based approach that allows companies and organizations to detect opportunities for improving the quality of their products or services through sentiment analysis. This approach is based on convolutional neural network (CNN) and word2vec. To determine the effectiveness of this approach for classifying tweets, we conducted experiments with different sizes of a Twitter corpus composed of 100000 tweets. We obtained encouraging results with a precision of 88.7%, a recall of 88.7%, and an F-measure of 88.7% considering the complete dataset.

1. Introduction

Nowadays, there is a lot of online opinions. This information is important for users because it helps them to make decisions about buying a product, voting in a political election, and choosing a travel destination, among other subjects. This information is also important for organizations since it helps them to know the general opinion about their products, the sales forecast, and the customer satisfaction in real time. Based on this information, companies can identify opportunities for improving the quality of their products or services.

A good example that demonstrates the importance of the opinions is a t-shirt of Zara clothing store which received negative opinions because it looked like the clothes used in the Holocaust. In these situations, companies must act quickly and solve the problem to avoid these opinions affecting their reputation. In this sense, to know the public opinion in real time is very important. Twitter is a social network, where users share information on almost everything in real

time. Therefore, companies consider this social network as a rich source of information that allows knowing the general opinion about their products and services, among others [1]. However, analyzing and processing all these opinions require much time and effort for the humans. On these grounds, a technology that processes automatically this information has arisen. This technology is known as sentiment analysis or opinion mining.

Sentiment analysis has been defined by several authors. However the definition most used in the research community is the proposed by Liu [2], who defined it as follows: "Sentiment analysis is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes."

In the last years, several approaches have been proposed for sentiment analysis. Most of these approaches are based on two main techniques, semantic orientation and machine learning. Although good results were obtained for both

techniques, several works in the literature have demonstrated that machine learning obtained better results. However, in more recent years a new technique known as deep learning has captured the attention of researchers because it has significantly outperformed traditional methods [3, 4]. Most of the deep-learning-based approaches for sentiment analysis are based on the English language. Hence, we propose a deep-learning-based approach for sentiment analysis of tweets in Spanish. Spanish is the third language most used on the Internet (http://www.internetworldstats.com/stats7.htm). Therefore, we consider that new approaches for sentiment analysis in the Spanish language are necessary.

The remainder of the paper is structured as follows. Section 2 presents a review of the literature about sentiment analysis and deep learning. Section 3 described the proposed approach. The experiments and results are presented in Section 4. Finally, Section 5 presents conclusions and future work.

2. Related Works

In the literature, several authors have proposed approaches for the sentiment analysis. These works have used two main techniques, semantic orientation and machine learning. With respect to the first technique, approaches use sentiment lexicons to determine the polarity. SentiWordNet is the most used lexicon in the literature [5, 6]. This lexicon is based on WordNet and it contains multiple senses of a word. Also, it provides a positive, objective, and negative value for each sense. Several works using this technique have obtained promising results; however, some other works have not obtained good results due to two main reasons: (1) sentiment lexicons mainly are based on English, which forces researchers to translate the English lexicons to the target language and (2) a word can have different senses depending on the domain where they are used.

Regarding the machine learning approach, authors use classification algorithms such as Support Vector Machines (SVM) [7–11], Bayesian Networks (BayesNet) [12], and decision trees (J48) [10], among others. For this technique, two data sets are necessary, a training set and an evaluation set. The training set is used for the algorithm to learn from features of the domain. Meanwhile, the evaluation set is used to validate the built model from the training set. The performance of the machine learning technique depends on the effectiveness of the selected method for feature extraction. Among the most used methods are bag of words [13], TF-IDF [14], n-grams (unigrams, bigrams, and trigrams) [11, 15], features based on POS tagging [16], and features based on dependency rules [17].

However, most recent works are based on deep learning techniques. For instance, Dos Santos and Gatti [18] proposed an approach to sentiment analysis of short texts. The approach is based on convolutional neural network, which is applied on two corpora, movies reviews (Stanford Sentiment Tree-bank) and Twitter messages (Stanford Twitter Sentiment corpus). Araque et al. [19] introduced an approach based on deep learning for sentiment classification. The

authors used a word embeddings model and a machine learning algorithm. To evaluate the performance of the proposed approach, the authors used six corpora publicly available of Twitter and movies reviews. Hu et al. [20] proposed a framework based on neural network for sentiment analysis. This framework is composed of two main phases. Firstly, feature vectors are obtained through linguistic and domain knowledge. Secondly, a Deep Neural Network is designed. Also, the authors evaluated their approach on three datasets (electronic products, movies reviews, and hotels reviews). Tang et al. [21] built a supervised learning framework. The authors combined sentiment features and features related to emoticons, negation, punctuation, cluster, and n-grams. Then, they trained a classifier by using a benchmark corpus provided in SemEval 2013. Ruder et al. [22] proposed an approach to aspect-based sentiment analysis. The authors used a convolutional neural network (CNN) for aspect extraction and sentiment analysis. The proposal was evaluated in several domains such as restaurants, hotels, laptops, phones, and cameras. Severyn and Moschitti [4] introduced a deep learning model which is applied to two tasks of SemEval 2015, namely, message-level and phrase-level of Twitter sentiment analysis. Sun et al. [23] proposed a sentiment analysis approach for Chinese microblog with a Deep Neural Network model. The proposed method extracted features to obtain semantic and information of words. Finally, three models, SVM, Naïve Bayes, and Deep Neural Network, are selected to prove the effectiveness of the method. Finally, Poria et al. [24] presented an approach to aspect extraction for sentiment analysis by using a deep learning technique. Also, the authors obtained a set of linguistic patterns to combine them with neural networks.

On the other hand, the approaches for the sentiment analysis are mainly focused on the analysis of the opinions of blogs, forums, and travel and sales websites. However, recently a more special interest has arisen on social networks such as Twitter because a lot of information from different topics can be extracted for its analysis. Among the most studied domains in the sentiment analysis area are movies, technological products, tourism, and health. Finally, regarding the language, most of them are based on English language and only one is based on the Chinese language.

Next section describes the deep-learning-based approach for sentiment analysis proposed in this work. More specifically, this section describes the architecture of our proposal as well as the relations among all its components.

3. Approach

The sentiment classification approach presented in this work is divided into three main modules: (1) preprocessing module, (2) word embeddings, and (3) CNN model. Figure 1 shows the workflow of the system. Firstly, the tokenization and normalization of the text are carried out. Secondly, word2vec is used to obtain the feature vectors. The last step consists in training a convolutional neural network to classify tweets as positives or negatives. A detailed description of these modules is provided in the following sections.

Parece q tenías razón *@bufalo58* y tendré q cambiarme a iPhone, xq el servicio técnico de *@SamsungChile* no va a reparar mi celu **#ChaoSamsumg**	*Mentions and replies* **Hashtags** <u>URL</u>

<div align="center">Box 1</div>
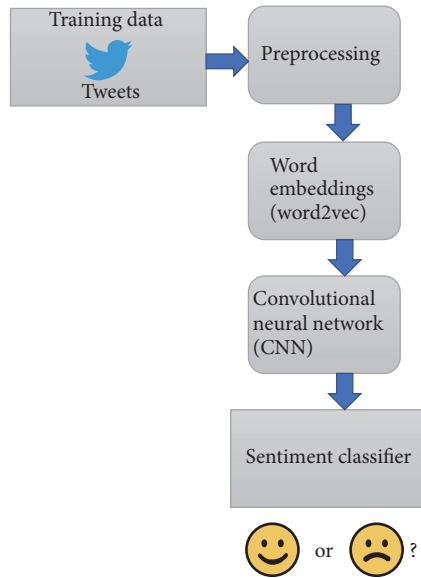

FIGURE 1: Our proposed approach.

FIGURE 2: Example of a tweet.

3.1. Preprocessing Module. The first step of the method proposed consists in the preprocessing of the tweets. Twitter is a social network where users use informal language due to the limitation of 140 characters. Therefore, there are several issues such as spelling errors, slang words, abbreviations, and replication of characters, among others, that must be addressed before detecting the polarity of a tweet. Figure 2 presents a tweet with some of these issues. In order to deal with this problem, we adopted the approach presented in [25] for the tweets processing.

The first phase of the preprocessing module consists in the tokenization process. In this process, the text is divided in tokens, which can be words or punctuation marks. To perform this process, the Twokenize (http://www.cs.cmu.edu/~ark/TweetNLP/) tool was used. This tool is oriented to

Twitter and allows identifying items of Twitter like hashtags, mentions and replies, and URLs, among others.

The second phase of this module consists in the normalization of the text. Firstly, items identified by Twokenize are removed because they do not provide important information for the detection of polarity. Next, each item removed from tweets is described.

(1) Mentions and replies to users: these items are represented with @.

(2) URLs: all items start with http://

(3) Hashtags: in this case, the character # is only removed due to the rest of the text representing an important part to be analyzed.

For instance, let us consider the tweet presented in Figure 2: "Parece q tenías razón @bufalo58 y tendré q cambiarme a iPhone, xq el servicio técnico de @SamsungChile no va a reparar mi celu #ChaoSamsung—It looks like you were right @bufalo58 and I will have to switch to iPhone because the @SamsungChile technical service is not going to repair my cell phone #ChaoSamsung." In this step, Twokenize detects two mentions and one Hashtag. Then, the module removes the mentions ("@bufalo58" and "@SamsumgChile") and the character "#" of "#ChaoSamsumg" (see Box 1).

Secondly, hashtags (strings that contain one or more words) are split based on capital letters. Considering the example presented above, #ChaoSamsung is split into two words "Chao" and "Samsung."

Thirdly, abbreviations and shorthand notations are extended. To this aim, we used the NetLingo (http://www.netlingo.com) dictionary. For example, "que" instead of "q," "por que" instead of "xq," and "celular" instead of "celu." Finally, Hunspell (http://hunspell.github.io) dictionary is used to correct spelling errors.

3.2. Word Embeddings. In this approach, we use word2vec for learning word embeddings. This tool implements the continuous bag-of-words model (CBOW) and skip-gram model for computing vector representations of words [26]. Word embeddings represent an important part in CNN architecture due to the fact that it allows obtaining syntactic and semantic information from the tweets, which is very important for sentiment classification.

3.3. CNN Model. We use a deep convolutional neural network for classification of tweets into positive and negative classes. The CNN (convolutional neural network) architecture requires concatenated word vectors of the text as input. Regarding the implementation of this model, Tensorflow (https://www.tensorflow.org) was used.
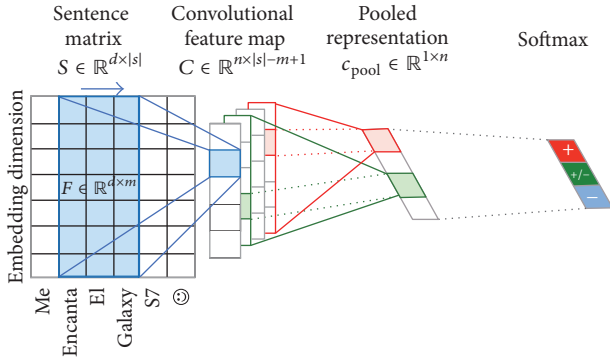
FIGURE 3: Learning model for sentiment classification [4].

Next, Figure 3 shows the architecture of a convolutional neural network used for sentiment classification [4].

4. Experiments

4.1. Data. The main objective of this approach is to detect important information about products and services that allows companies and organizations to improve them. Therefore, our approach requires a corpus related to products and services. Although several corpora have been provided in the literature, there is a lack of corpora for Spanish. In this sense, we have obtained a corpus from Twitter in Spanish. The process for collecting this corpus is described below.

(1) Tweets were collected by using Twitter4J (http://twitter4j.org/) library. To obtain relevant tweets, a set of keywords related to technological products were defined.

(2) Duplicated tweets, retweets, tweets in other languages, and tweets that contain only URLs were removed.

(3) We obtained a total of 70000 positive tweets and 63000 negative tweets.

(4) Finally, we selected only 50000 positive tweets and 50000 negative tweets, which were manually analyzed to obtain those relevant to our study.

This corpus is not available publicly because according to the Twitter privacy policy it is not possible to share the content of the tweets. Next, two examples from the corpus collected are presented. Figure 4 shows a positive tweet "Una excelente característica del iPhone 7 #JumboMobile @tiendasjumboco es su resistencia al agua—An excellent feature of the iPhone 7 #JumboMobile @tiendasjumboco is its water resistance," while Figure 5 shows an example of a negative tweet "lo que quise dar a entender es que n me salio bueno ni el cargador ni el iPhone pq se me rompieron los dos—What I meant to say was, both the charger and iPhone were not good because the two broke."

Table 1 shows the distribution of our corpus. As can be seen, 40000 positive and negative tweets were used to train the classifier and 10000 tweets positive and negative were used to test the model built.
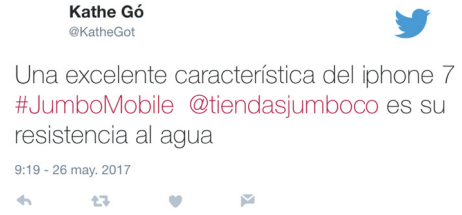

FIGURE 4: Positive tweet.

FIGURE 5: Negative tweet.

TABLE 1: Distribution of the corpus.

	Positive	Negative	Total
Train	40000	40000	80000
Test	10000	10000	20000

4.2. Evaluation and Results. Aiming to measure the performance of our proposed approach, we have used well-known metrics: precision, recall, and F-measure. Precision (1) represents the proportion of predicted positive cases that are real positives. On the other hand, recall (2) is the proportion of actual positive cases that were correctly predicted as such. F-measure (3) is the harmonic mean of precision and recall [27].

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}, \quad (1)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}, \quad (2)$$

$$F\text{-measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3)$$

Also, we used the macro precision (4), macro recall (5), and macro F-measure (6) metrics due to the fact that the polarity detection is a multiclass problem.

$$\text{Macro precision} = \frac{\sum_{i=1}^{|C|} \text{Precision}}{|C|}, \quad (4)$$

$$\text{Macro Recall} = \frac{\sum_{i=1}^{|C|} \text{Recall}}{|C|}, \quad (5)$$

$$\text{Macro } F\text{-measure} = \frac{\sum_{i=1}^{|C|} F\text{-measure}}{|C|}. \quad (6)$$

Table 2 shows that our approach obtained encouraging results with a precision of 88.5%, a recall of 88.8%, and an F-measure

TABLE 2: Results obtained with CNN.

	Precision	Recall	F-measure
Positive	0.885	0.888	0.887
Negative	0.888	0.884	0.886
Macro	**0.887**	**0.887**	**0.887**

TABLE 3: Results of traditional model and Deep model.

	CNN	SVM	NB
5000	0.785	0.675	0.680
10000	0.793	0.723	0.712
20000	0.843	0.805	0.802
40000	0.844	0.808	0.779
60000	0.852	0.811	0.771
80000	0.858	0.833	0.777
100000	0.887	0.837	0.779

FIGURE 6: Comparison of SVM, NB, and CNN for different sizes of subsets for the Twitter dataset.

of 88.7% for the positive class, and a precision of 88.8%, a recall of 88.4%, and an F-measure of 88.6% for the negative class.

4.3. Comparison with Traditional Learning Methods. In this work, different classification algorithms were compared with the same feature vector, namely, SVM, NB, and CNN (see Table 3). For a fair comparison, the default parameters were used for each algorithm without carrying out an additional tuning process. This analysis was carried out in order to study the effects of the proposed approach with a convolutional neural network. The algorithms were evaluated with several sizes of the corpus. Each subset is split into two datasets: (1) 80% of the data is used as a training set and (2) 20% of the data is used as a testing set.

As can be seen in Figure 6, traditional models show similar results. However, SVM provides better results than NB when the size of data increases. On the other hand, results also indicate that convolutional neural network obtained better results that traditional models (SVM and NB) with the different subsets of the Twitter corpus. These results confirm that deep learning techniques outperformed traditional methods of machine learning for sentiment analysis.

It is important to mention that we did not carry out a comparison of our results with those reported in related works because there is a lack of deep learning approaches for sentiment analysis in Spanish.

5. Conclusions and Future Work

In this work, we presented an approach for Twitter sentiment analysis. The main objective of this proposal was providing the basis to know customer satisfaction and identify opportunities for improvement of products and services. The proposal is based on a deep learning model to build a classifier for sentiment detection. Our approach obtained encouraging results, with a precision, recall, and F-measure of 88.7%. The results also show that CNN outperformed traditional models such as SVM and NB.
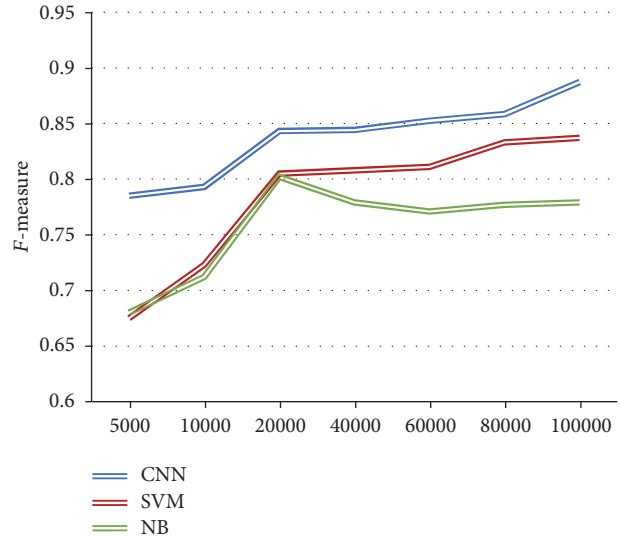
As future work, we are considering exploring other neural network models such as Recursive Neural Tensor Networks (RNTN), Recurrent Neural Networks (RNN), and Long Short Term Memory (LSTM). Also, we plan to evaluate other word embedding features as those presented in [21]. Finally, we have considered applying our approach to other languages such as English, French, and Arabic.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
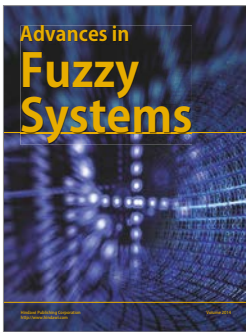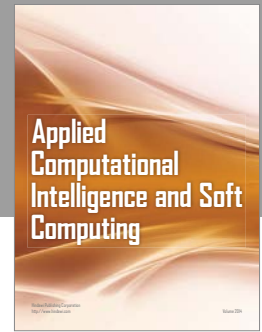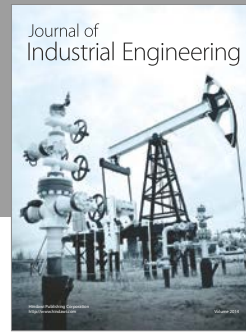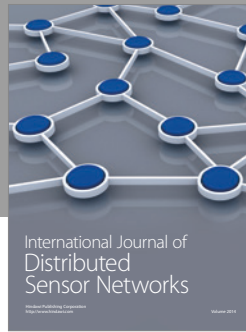
Acknowledgments

References

[1] A. Das, S. Gollapudi, and K. Munagala, "Modeling opinion dynamics in social networks," in *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*, pp. 403–412, 2014.

[2] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[3] M. Cliche, "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs," Apr. 2017.

[4] A. Severyn and A. Moschitti, "UNITN: training deep convolutional neural network for twitter sentiment classification,"

in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval '15)*, pp. 464–469, 2015.

[5] I. Peñalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia et al., "Feature-based opinion mining through ontologies," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5995–6008, 2014.

[6] A. Montejo-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, "Ranked word net graph for sentiment polarity classification in Twitter," *Computer Speech & Language*, vol. 28, no. 1, pp. 93–107, 2014.

[7] P. Biyani, C. Caragea, P. Mitra et al., "Co-training over Domain-independent and Domain-dependent features for sentiment analysis of an online cancer support community," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*, pp. 413–417, ACM, Sydney, Australia, August 2013.

[8] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: an empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[9] N. U. Pannala, C. P. Nawarathna, J. T. K. Jayakody, L. Rupasinghe, and K. Krishnadeva, "Supervised learning based approach to aspect based sentiment analysis," in *Proceedings of the 16th IEEE International Conference on Computer and Information Technology (CIT '16)*, pp. 662–666, 2016.

[10] M. Venugopalan and D. Gupta, "Exploring sentiment analysis on twitter data," in *Proceedings of the 8th International Conference on Contemporary Computing (IC3 '15)*, pp. 241–247, August 2015.

[11] M. del Pilar Salas-Zárate, M. A. Paredes-Valverde, J. Limon-Romero, D. Tlapa, and Y. Baez-Lopez, "Sentiment classification of Spanish reviews: an approach based on feature selection and machine learning methods," *Journal of Universal Computer Science*, vol. 22, no. 5, pp. 691–708, 2016.

[12] M. Del Pilar Salas-Zárate, E. López-López, R. Valencia-García, N. Aussenac-Gilles, Á. Almela, and G. Alor-Hernández, "A study on LIWC categories for opinion mining in Spanish reviews," *Journal of Information Science*, vol. 40, no. 6, pp. 749–760, 2014.

[13] N. F. F. Da Silva, E. R. Hruschka, and E. R. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Decision Support Systems*, vol. 66, pp. 170–179, 2014.

[14] E. Fersini, E. Messina, and F. A. Pozzi, "Sentiment analysis: bayesian ensemble learning," *Decision Support Systems*, vol. 68, pp. 26–38, 2014.

[15] P. Smith and M. Lee, "Cross-discourse development of supervised sentiment analysis in the clinical domain," in *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 79–83, 2012.

[16] I. Habernal, T. Ptáček, and J. Steinberger, "Reprint of "supervised sentiment analysis in Czech social media"," *Information Processing and Management*, vol. 51, no. 4, pp. 532–546, 2015.

[17] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, and A. Hussain, "Concept level sentiment analysis using dependency-based semantic parsing: a novel approach," *Cognitive Computation*, vol. 7, no. 4, pp. 487–499, 2015.

[18] C. N. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of the 25th International Conference on Computational Linguistics (COLING '14)*, pp. 69–78, 2014.

[19] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Systems with Applications*, vol. 77, pp. 236–246, 2017.

[20] Z. Hu, J. Hu, W. Ding, and X. Zheng, "Review Sentiment Analysis Based on Deep Learning," in *Proceedings of the 12th IEEE International Conference on E-Business Engineering (ICEBE '15)*, pp. 87–94, 2015.

[21] D. Tang, F. Wei, B. Qin, T. Liu, and M. Zhou, "Coooolll: A Deep Learning System for Twitter Sentiment Classification," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval '14)*, pp. 208–212, Dublin, Ireland, 2014.

[22] S. Ruder, P. Ghaffari, and J. G. Breslin, "INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval '16)*, pp. 330–336, San Diego, CA, USA, June 2016.

[23] X. Sun, C. Li, and F. Ren, "Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features," *Neurocomputing*, vol. 210, pp. 227–236, 2016.

[24] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Systems*, vol. 108, pp. 42–49, 2016.

[25] M. d. Salas-Zárate, M. A. Paredes-Valverde, M. Á. Rodriguez-García, R. Valencia-García, and G. Alor-Hernández, "Automatic detection of satire in Twitter: A psycholinguistic-based approach," *Knowledge-Based Systems*, vol. 128, pp. 20–33, 2017.

[26] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS '13)*, pp. 3111–3119, December 2013.

[27] M. D. P. Salas-Zárate, R. Valencia-García, A. Ruiz-Martínez, and R. Colomo-Palacios, "Feature-based opinion mining in financial news: an ontology-driven approach," *Journal of Information Science*, 2016.