Chapter 5
Deep Learning for Twitter Sentiment Analysis: The Effect of Pre-trained Word Embedding

Akrivi Krouska, Christos Troussas, and Maria Virvou

Abstract Twitter is the most popular microblogging platform, with millions of users exchanging daily a huge volume of text messages, called "tweets". This has resulted in an enormous source of unstructured data, Big Data. Such a Big Data can be analyzed by companies or organizations with the purpose of extracting customer perspective about their products or services and monitoring marketing trends. Understanding automatically the opinions behind user-generated content, called "Big Data Analytics", is of great concern. Deep learning can be used to make discriminative tasks of Big Data Analytics easier and with higher performance. Deep learning is an aspect of machine learning which refers to an artificial neural network with multiple layers and has been extensively used to address Big Data challenges, like semantic indexing, data tagging and immediate information retrieval. Deep learning requires its input to be represented as word embeddings, i.e. as a real-value vector in a high-dimensional space. However, word embedding models need large corpuses for training and presenting a reliable word vector. Thus, there are a number of pre-trained word embeddings freely available to leverage. In effect, these are words and their corresponding *n*-dimensional word vectors, made by different research teams. In this work, we have made data analysis with huge numbers of tweets taken as big data and thereby classifying their polarity using a deep learning approach with four notable pre-trained word vectors, namely Google's Word2Vec, Stanford's Crawl GloVe, Stanford's Twitter GloVe, and Facebook's FastText. One major conclusion is that tweet classification using deep learning outperforms the baseline machine

A. Krouska · C. Troussas · M. Virvou (✉)
Software Engineering Laboratory, Department of Informatics, University of Piraeus, Piraeus, Greece
e-mail: mvirvou@unipi.gr

A. Krouska
e-mail: akrouska@unipi.gr

C. Troussas
e-mail: ctrouss@unipi.gr

111

learning algorithms. At the same time and with regard to pre-trained word embeddings, FastText provides more consistent results across datasets, while Twitter GloVe obtains very good accuracy rates despite its lower dimensionality.

Keywords Deep learning · Big data · Twitter · Sentiment analysis · Word embeddings · CNN

5.1 Introduction

The proliferation of Web 2.0 has led to a digital landscape where people are able to socialize, by expressing and sharing their thoughts and opinions on various issues in public, through a variety of means and applications. Indicative of this innovation of social interaction is microblogging, an online broadcast medium that allows users to post and share short messages with an audience online. Twitter is the most popular of such services with million users exchanging daily a huge volume of text messages, called tweets. This has resulted in an enormous source of unstructured data, Big Data, which can be integrated into the decision-making process, Big Data Analytics. Big data analytics is a complex process of examining large and varied data sets (Big Data) to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make better-informed business decisions [1]. Twitter is a valuable media for this process. This paper focuses on the analysis of opinions of tweets, called Twitter Sentiment Analysis.

Sentiment Analysis is the process aiming to detect sentiment content of a text unit in order to identify people's attitudes and opinions towards various topics [2]. Twitter, with nearly 600 million users and over 250 million messages per day, has become one of the largest and most dynamic datasets for sentiment analysis. Twitter Sentiment Analysis refers to the classification of tweets based on the emotion or polarity that the user intends to transmit. This information is extremely valuable to numerous circumstances where the decision making is crucial. For instance, the opinion and sentiment detection is useful in politics to forecast election outcomes or estimate the acceptance of politicians [3], in marketing for sales predictions, product recommendations and investors' choices [1], and in the educational context for incorporating learner's emotional state to the student model providing an affective learning environment [4].

Twitter Sentiment analysis is becoming increasingly important for social media mining, as it gives the access to valuable opinions of numerous participants on various business and social issues. Consequently, many researchers have focused on the study of applications and enhancements on sentiment analysis algorithms that provide more efficient and accurate results [5]. There are three main sentiment analysis approaches: a. machine learning-based, which uses classification technique to classify the text entity, b. lexicon-based, which uses sentiment dictionary with opinion words and weights determined the polarity, and c. hybrid, where different approaches are combined [4, 6, 7]. Experimental results have shown that machine

learning methods have higher precision, while lexicon-based methods are competitive in case the training dataset lacks quantity and quality, as they require few efforts in human-labeled document [5, 7]. On the other hand, combing the proper techniques foster better results, as the hybrid approaches could collectively exhibit the accuracy of a machine learning algorithms and the speed of lexicons [6]. However, recently a new approach has arisen from machine learning field, namely deep learning, overshadowing the aforementioned methods.

Deep learning is an aspect of machine learning which refers to an artificial neural network with multiple layers: an input, an output and at least one hidden [8]. The "deep" in deep learning refers to having more than one hidden layer. Replicating the human brain, neural networks consist of a large number of information processing units, called neurons, organized in layers, which work in unison. Deep learning is a hierarchical feature learning, where each layer learns to transform its input into a slightly more abstract and composite representation through a nonlinear processing, using a weight-based model with an activation function to each neuron in order to dictate the importance of the input value. Iterations continue until an acceptable rate of accuracy reached at output data.

Deep learning can be used to extract incredible information that buried in a Big Data [1]. It has been extensively applied in artificial intelligence field, like computer vision, transfer learning, semantic parsing, and natural language processing. Thus, for more accurate sentiment analysis, many researchers have implemented deferent models of this approach, namely CNN (convolutional neural networks), DNN (deep neural networks), RNN (recurrent neural networks) and DBN (deep belief networks) [8]. The results show that deep learning is very beneficial in the sentiment classification. Except the higher performance, there is no need for carefully optimized hand-crafted features and feature engineering, one of the most time-consuming parts of machine learning practice. Big Data challenges, like semantic indexing, data tagging and immediate information retrieval, can be addressed better using Deep learning. However, it requires large data sets and is extremely computationally expensive to train.

In this study, a convolutional neural network was applied on three well-known Twitter datasets, namely Obama-McCain Debate (OMD), Health Care Reform (HCR) and Stanford Twitter Sentiment Gold Standard (STS-Gold, for sentiment classification, using different pre-trained word embeddings. Word Embeddings is one of the most useful deep learning methods used for constructing vector representations of words and documents. Their novelty is the ability to capture the syntactic and semantic relations among words. The most successful deep learning methods of word embeddings are Word2Vec [9, 10], GloVe [11] and FastText [12]. Many researchers have used these methods in their sentiment analysis experiments [13, 14]. Despite their effectiveness, one of their limitations is the need of large corpuses for training and presenting an acceptable word vector. Thus, researches have to use pre-trained word vectors, due to the small size of some datasets. In this research, we experiment on four notable pre-trained word embeddings, namely Google's Word2Vec, Stanford's Crawl GloVe, Stanford's Twitter GloVe, and Facebook's FastText, and examine their effect on the performance of sentiment classification.

The structure of this paper is as follows. First, we present the related work in Twitter sentiment analysis using deep learning techniques. Following, we present the evaluation procedure of this research, describing the datasets, the pre-trained word embeddings and the deep learning algorithm used. Section 5.4 deals with the comparative analysis and discussion on the experiment results. Finally, we present our conclusions and future work.

5.2 Related Work

The need for analyzing the big data originated from the proliferation of social media and classifying sentiment efficiently lead many researchers in the adoption of deep learning models. The models used vary in the dimensions of the datasets employed, the deep learning techniques applied regarding the word embeddings and algorithm, and the purpose served. This section briefly describes representative studies related to Twitter sentiment analysis using deep learning approaches, and tables them based on the aforementioned dimensions.

In [13], the authors experiment with deep learning models along with modern training strategies in order to achieve better sentiment classifier for tweets. The proposed model was an ensemble of 10 CNNs and 10 LSTMs together through soft voting. The models ensembled were initialized with different random weights and used different number of epochs, filter sizes and embedding pre-training algorithms, i.e. Word2Vec or FastText. The GloVe variation is excluded from ensembled model, as it gave a lower score than both the other two embeddings.

In [15], the authors propose a three-step process to train their deep learning model for predicting polarities at both message and phrase levels: i. word embeddings are initialized using Word2Vec neural language model, which is trained on a large unsupervised tweet dataset; ii. a CNN is used to further refine the embeddings on a large distant supervised dataset; iii. the word embeddings and other parameters of the network obtained at the previous stage are used to initialize the network with the same architecture, which is then trained on a supervised dataset from Semeval-2015.

In [16], the authors develop a Bidirectional Long Short-Term Memory (BLSTM) RNN, which combines 2 RNNs: a forward RNN where the sequence is processed from left to right, and a reverse RNN where the sequence is processed from right to left. The average of both RNNs outputs for each token is used to compute the model's final label for adverse drug reaction (ADR) identification. For pre-trained word embeddings, they utilized the Word2Vec 400 M Twitter model (by Frederic Godin),[1] which is a set of 400-dimensional word embeddings trained using the skip-gram algorithm on more than 400 million tweets.

In [17], the authors propose a new classification method based on deep learning algorithms to address issues on Twitter spam detection. They firstly collected a 10-day

---

[1] https://github.com/loretoparisi/word2vec-twitter.

real tweets corpus and applied Word2Vec to pre-process it, instead of feature extraction. Afterwards, they implemented a binary detection using MLP and compared their approach to other techniques, such as traditional text-based methods and feature-based methods supported by machine learning.

In [18], the authors apply convolution algorithm on Twitter sentiment analysis to train deep neural network, in order to improve the accuracy and analysis speed. Firstly, they used GloVe model to implement unsupervised learning of word-level embeddings on a 20 billion twitter corpus. Afterwards, they combined this word representation with the prior polarity score feature and sentiment feature set, and used it as input in a deep convolution neural network to train and predict the sentiment classification labels of five Twitter datasets. Their method was compared to baseline approaches regarding preprocessing techniques and classification algorithms used.

In [19], the authors present an effective deep neural architecture for performing language-agnostic sentiment analysis over tweets in four languages. The proposed model is a convolutional network with character-level embedding, which is designed for solving the inherent problem of word-embedding and n-gram based approaches. This model was compared to other CNN and LSTM models with different embedding, and a traditional SVM-based approach.

In [14], the authors, firstly, construct a tweet processor using semantic rules, and afterwards, train character embedding DeepCNN to produce feature maps capturing the morphological and shape information of a word. Moreover, they integrated global fixed-size character feature vectors and word-level embedding for Bi-LSTM. As pretrained word vectors, they used the public available Word2Vec with 300 dimensionality and Twitter GloVe of Stanford with 200 dimensionality. The results showed that the use of character embedding through a DeepCNN to enhance information for word embedding built on top of Word2Vec or GloVe improves classification accuracy in Twitter sentiment analysis.

In [20], the authors apply deep learning techniques to classify sentiment of Thai Twitter data. They used Word2Vec to train initial word vectors for LSTM and Dynamic CNN methods. Afterwards, they studied the effect of deep neural network parameters in order to find the best settings, and compared these two methods to other classic techniques. Finally, they demonstrated that the sequence of words influences sentiment analysis.

Feeding the network with the proper word embeddings is a key factor for achieving an effective and punctual sentiment analysis. Therefore, this article is concentrated on the evaluation of the most popular pre-trained word embeddings, namely Word2Vec, Crawl GloVe, Twitter GloVe, FastText. The experiments were conducted using CNN method in the context of sentiment analysis on three Twitter datasets.

Table 5.1 illustrates the analysis of deep learning researches described in this section, analyzing them regarding the datasets used, the deep learning techniques applied, and the scope served. It is worth noting that the authors mainly used the well-known word vectors Word2Vec and GloVe for training their network and CNN as base algorithm. Moreover, in the vast majority of cases, the purpose is to present a comparative analysis of deep learning approaches reporting their performance metrics.

Table 5.1 Analysis of deep learning researches

| Research	Tweets dataset	Num. of tweets	Deep learning techniques		Scope
			Word embeddings	Classification	
Cliche [13]	SemEval-2017	12.379 test set	Word2Vec, FastText, GloVe	Ensemble of CNN and LSTM	Algorithms evaluation
Severyn and Moschitti [15]	Twitter'15	5.482 test set	Word2Vec	CNN	Algorithm evaluation
Cocos [16]	Combination of Adverse Drug Reaction (ADR) dataset and Attention Deficit Hyperactivity Disorder (ADHD) dataset	844 (634 training set–210 test set)	Word2Vec 400 M Twitter model (by Frederic Godin)[a]	BLSTM	Adverse Drug Reaction (ADR) detection
Wu et al. [17]	Real-life 10-day ground-truth dataset	1.376.206 spam tweets + 673.836 non-spam	Word2Vec	MLP	Spam detection
Jianqiang et al. [18]	STSTd, SE2014, STSGd, SSTd, SED	359, 5.892, 2.034, 3.326, 2.648 correspondingly	GloVe	DeepCNN	Algorithm evaluation
Wehrmann et al. [19]	Tweets from four European languages	128.189 (89.732 training, 12.819 validation, 25.638 test)	Word and character-level embedding	CNN, LSTM	Language-agnostic translation
Nguyen and Nguyen [14]	STS Corpus, Sanders, HCR	80 K train + 16 K dev + 359 test, 991 train + 110 dev + 122 test, 621 train + 636 dev + 665 test correspondingly	Word2Vec, Twitter GloVe	DeepCNN, BLSTM	Algorithm evaluation
Vateekul and Koomsubha [20]	Thai Twitter Data	3.813.173	Word2Vec	Dynamic CNN, LSTM	Algorithm evaluation
Ours	OMD, HCR, STS-Gold	1.904, 1.922, 2.034 correspondingly	Word2Vec, Crawl GloVe, Twitter GloVe, FastText	CNN	Word embeddings evaluation

[a]https://github.com/loretoparisi/word2vec-twitter

5.3   Evaluation Procedure

The scope of the current research is to investigate the effect of the top pre-trained word embeddings obtained by unsupervised learning on large text corpora on training a deep convolutional neural network for Twitter sentiment analysis. Figure 5.1 illustrates the steps followed.

Firstly, word embeddings is used to transform tweets into its corresponding vectors to build up the sentences vectors. Then, the generated vectors are divided into training and test set. The training set is used as input to CNN in order to be trained and make predictions on the test set. Finally, the experiments' outcomes have been tabulated and a descriptive analysis has been conducted, compared CNN results with baseline machine learning approaches. All the preprocessing settings and the classification were employed in Weka data mining package.

5.3.1   Datasets

The experiments described in this paper were performed in three well-known and freely-available on the Web Twitter datasets. The reason they have been chosen is that they consist of a significant volume of tweets, created by reputable universities for academic scope, and they have been used in various researches. STS-Gold dataset [21] consists of random tweets with no particular topic focus, whereas OMD [22] and HCR [23] include tweets on specific domains. In particular, OMD dataset contains tweets related with Obama-McCain Debate and HCR dataset focuses on Health Care Reform. The statistics of the datasets are shown in Table 5.2 and examples of tweets are given in Table 5.3.
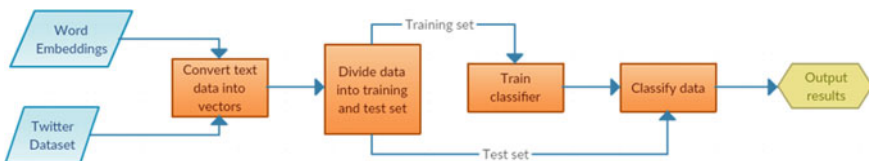
Fig. 5.1   Twitter classification using word embeddings to train CNN

Table 5.2   Statistics of the three Twitter datasets used

Dataset	Tweets	Positive	Negative
Obama-McCain Debate (OMD)	1904	709	1195
Health Care Reform (HCR)	1922	541	1381
Stanford Sentiment Gold Standard (STS-Gold)	2034	632	1402

Table 5.3 Examples of tweets

Dataset	Tweet	Polarity
OMD	the debate was a draw. I thought Mccain did come out slightly on top because he sounded more educated on the issues. good closing. #debate08	Positive
	Good debate. Job well done to both candidates! #current, #debate08, #tweetdebate	
	I am disappointed. #tweetdebate	Negative
	#current #tweetdebate It's sad that you can watch this and think that Mc has been the only one saying anything right tonight!	
HCR	Healthcare is not a privilege, it is a right! That is why I support healthcare reform. #hcr	Positive
	Maybe you healthy and happy with your healthcare plan and that's ok. But don't stand in the way of the people who are sick and in need #HCR	
	You know its a bad bill when you have to payoff your own party #tcot #tweetcongress #killthebill #HCR #handsoff	Negative
	Retweet this if you are against a government takeover of health care #hcr #tcot #gop #txgop #handsoff #codered	
STS-Gold	I love my new tiny cute little iPod! Thank you @Santino_gq! Xoxoxox!!!	Positive
	sooo excited….i just bought tickets to Taylor Swift. ahhhhhh	
	So disapointed Taylor Swift doesn't have a Twitter	Negative
	Trying to set up Xbox Live and failing tremendously…brain ache	

5.3.2 Data Preprocessing

Due to the short length of tweets and their informal aspect, tweets are usually consisted of a variety of noise and language irregularities. This noise and unstructured sentences will affect the performance of sentiment classification [24]. Thus, before feature selection, a stage of preprocessing is essential to be involved in the classification task. The preprocessing includes:

- Stemming (reducing inflected words to their word stem, base or root form).
- Remove all numbers, punctuation symbols and some special symbols.
- Force lower case for all tokens.
- Remove stopwords (referred to the most common words in a language).
- Replace the emoticons and emoji with their textual form using an emoticon dictionary.
- Tokenize the sentence.

In our experiments, we use Wikipedia emoticons list,[2] Weka Stemmer, Rainbow stopwords list, and Tweet-NLP.

---

[2]http://en.wikipedia.org/wiki/List_of_emoticons.

5.3.3 Pre-trained Word Embeddings

Word embeddings are a set of natural language processing techniques where individual words are mapped to a real-value vector in a high-dimensional space [9]. The vectors are learned in such a way that words that have similar meanings will have similar representation in the vector space. This is a more semantic representation for text than more classical methods like bag-of-words, where relationships between words or tokens are ignored, or forced in bigram and trigram approaches [10]. This numeric representation is essential in sentiment analysis, as many machine learning algorithms and deep learning approaches are incapable of processing strings in their raw form and require numeric values as input to perform their tasks. Thus, there are a variety of word embeddings methods which can be applied on any corpus in order to build a vocabulary with hundred of dimensions word vectors, capturing a great deal of semantic patters.

In 2013, Word2Vec [9, 10] has been proposed by Mikolov et al., which has now become the mainstream of word embedding. Word2Vec employs a two-level neural network, where Huffman techniques is used as hierarchical softmax to allocate codes to frequent words. The model is trained through stochastic gradient descent and the gradient is achieved by backpropagation. Moreover, optimal vectors are obtained for each word by CBOW or Skip-gram.

In 2014, Pennington et al. introduced GloVe [11], which become also very popular. GloVe is essentially a log-bilinear model with a weighted least-squares objective. The model is based on the fact that ratios of word-word co-occurrence probabilities have the potential for encoding some form of meaning.

In 2016, based on Word2Vec, Bojanowski et al. proposed FastText [12], which can handle subword units and has fast computing. FastText extends Word2Vec by introducing subword modeling. Instead of feeding individual words into the neural network, it represents each word as a bag of character n-gram. The embedding vector for a word is the sum of all its n-grams.

Word embedding is a demanding process, requiring large corpuses for training and presenting an acceptable word vector. Thus, rather than training the word vectors from scratch, it can be used pre-trained word embeddings, public available in Internet, for sentiment analysis through deep learning architectures. In this paper, we use four well-known pre-trained word vectors, namely Google's Word2Vec,[3] Stanford's Crawl GloVe,[4] Stanford's Twitter GloVe, and Facebook's FastText,[5] with the intention of evaluating their effect in sentiment analysis. A briefly description of these vectors is following.

Google's Word2Vec trained on part of Google News dataset, about 100 billion words. The model contains 300-dimensional vectors for 3 million words and phrases. Crawl GloVe was trained on a Common Crawl dataset of 42 billion tokens (words), providing a vocabulary of 2 million words with an embedding vector size of 300

---

[3]https://code.google.com/archive/p/word2vec/.

[4]https://nlp.stanford.edu/projects/glove/.

[5]https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md.

Table 5.4 Statistics of pre-trained word embeddings

Pre-trained word vector	Dataset	Dimension	Corpora	Vocabulary
Word2Vec	Google news	300	100 billion	3 million
Crawl GloVe	Common crawl	300	42 billion	2 million
Twitter GloVe	Twitter	200	27 billion	2 million
FastText	Wikipedia 2017, UMBC webbase corpus and statmt.org news	300	16 billions	1 million

dimensions. On the other hand, Twitter GloVe was trained on a dataset of 2 billion tweets with 27 billion tokens. This representation has a vocabulary of 2 million words and embedding vector sizes, including 50, 100 and 200 dimensions. Finally, FastText consists of 1 million-word vectors of 300 dimensions, trained on Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset (16 billions of tokens) (Table 5.4).

5.3.4 Deep Learning

In this paper, we use a deep Convolutional Neural Network (CNN) for tweets classification. CNN is a class of deep learning approaches which use a variation of multilayer perceptrons designed to require minimal preprocessing [25]. First, the tweet is tokenized and transformed into a list of word embeddings. Afterwards, the created sentence matrix is combined using multiple filters with variable window size. Thus, local sentiment feature vectors are generated for each possible word window size. Then, the feature maps activate a 1-max-pooling layer via a non-linear activation function. Finally, this pooling layer is densely connected to the output layer using softmax activation to generate probability value of sentiment classification, and optional dropout regularization to prevent over-fitting. The architecture of the convolutional neural network used for sentiment classification is shown on Fig. 5.2.

5.4 Comparative Analysis and Discussion

In this paper, we experimented on CNN for the sentiment analysis of three well-known dataset, using four different word embeddings. The scope of the paper is to evaluate the effect of pre-trained word embeddings, namely Word2Vec, Crawl GloVe, Twitter GloVe, and FastText, on the performance of Twitter classification. For all datasets, the same preprocessing steps were applied. Many experiments were made in order to define the best settings for CNN. Thus, we set a mini-batch size of 50. Learning rate technique was applied for stochastic gradient descent with a maximum
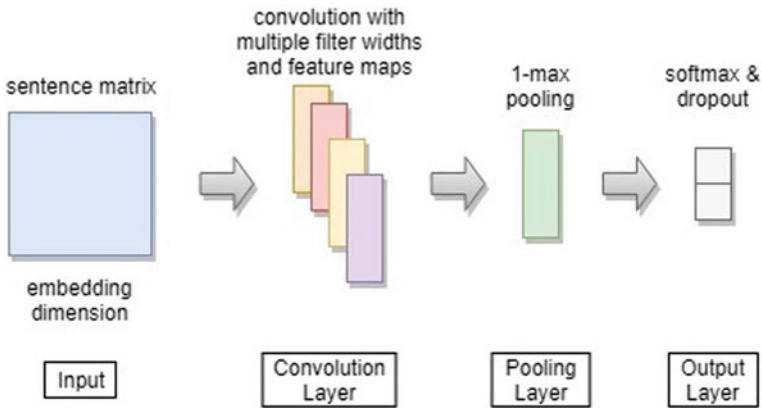
Fig. 5.2 Architecture of Convolutional Neural Network (CNN) used

of 100 training epochs. We tested with various filter windows and concluded to set the filter windows to 5. Moreover, we use the activation function ReLu for the convolution layer and SoftMax for the output one. The experimental evaluation metric is the accuracy in the classification of positive and negative tweets, and the best CNN results are compared to the performance of baseline classifiers emerged from other previous studies [5].

Table 5.5 illustrates the percentage of accuracy achieved by CNN regarding the different word embeddings used. We observe that using FastText gives better performance than other approaches in the majority of experiments. Moreover, Twitter Glove performs satisfactory, despite having lower dimensionality than other word vectors and training on a considerably smaller dataset. Its good accuracy rates may be due

Table 5.5 CNN accuracy using different word vectors on Twitter datasets

Dataset	Word embeddings	CNN accuracy
OMD	Word2Vec	81.98
	Crawl GloVe	83.59
	Twitter GloVe	82.79
	FastText	**83.65**
HCR	Word2Vec	79.94
	Crawl GloVe	80.45
	Twitter GloVe	80.39
	FastText	**81.33**
STS-Gold	Word2Vec	82.69
	Crawl GloVe	84.12
	Twitter GloVe	**84.23**
	FastText	82.57
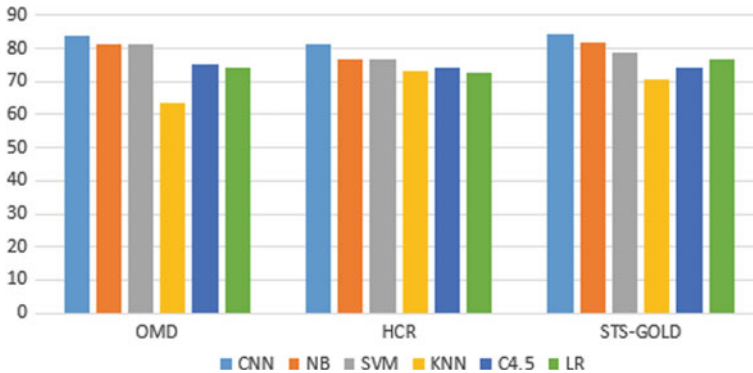
Fig. 5.3 Comparison of accuracy values between our best CNN version and baseline approaches

to the fact that the corpora used by this model originated from Twitter. On the other hand, Word2Vec seems not to be as effective as others models.

We further compare our method to five representative machine learning algorithms, namely Naïve Bayes, Support Vector Machine, k- Nearest Neighbor, Logistic Regression and C4.5. In [5], these classifiers have been evaluated using the same datasets as the current research, and their performance has been tabled. We use the best accuracy rates of CNN version for each dataset, in order to compare with the baseline algorithms. As represented in Fig. 5.3, the deep learning technique outperforms the other machine learning methods over all the datasets.

5.5 Conclusion and Future Work

Deep Learning and Big Data analytics are, nowadays, two burning issues of data science. Big Data analytics is important for organizations that need to extract information from huge amounts of data collected through social networking services. Such a social networking service is a microblogging service, with Twitter being its most representative platform. Through Twitter, millions of users exchange public messages daily, namely "tweets", expressing their opinion and feelings towards various issues. This makes Twitter one of the largest and most dynamic Big Data sets for data mining and sentiment analysis. Deep learning is a valuable tool for Big Data analytics. Deep Learning models have achieved remarkable results in a diversity of fields, including natural language processing and classification. Thus, in this paper, we employ a deep learning approach for Twitter sentiment classification, with the intention of evaluating different settings of word embeddings.

The development and use of word embeddings is an essential task in the classification process, as deep learning techniques require numeric values as input, and feeding the network with the proper input can boost the performance of algorithms. In this work, we examined four well-known pre-trained word embeddings, namely

Word2Vec, Crawl GloVe, Twitter Glove and FastText, that have been developed by reliable research teams and have been used extensively in deep learning projects. The results show that FastText provides more consistent results across datasets. Moreover, Twitter GloVe obtains very good accuracy rates despite its lower dimensionality.

Part of our future work is to study word embeddings trained from scratch on Twitter data using different parameters, as the training model, dimensional size, relevant vocabulary to test dataset etc., and evaluate their effect on classification performance. Moreover, another interesting research venue is concerned with the comparative analysis of different deep learning approaches regarding the algorithm structure, layers, filters, functions etc., in order to identify the proper network settings.

References

1. S. Sohangir, D. Wang, A. Pomeranets, T.M. Khoshgoftaar, Big Data: deep learning for financial sentiment analysis. J. Big Data **5**(1) 2018
2. V. Sahayak, V. Shete, A. Pathan, Sentiment analysis on twitter data. Int. J. Innov. Res. Adv. Eng. (IJIRAE) **2**(1), 178–183 (2015)
3. K.C. Tsai, L.L. Wang, Z. Han, Caching for mobile social networks with deep learning: twitter analysis for 2016 US election. IEEE Trans. Netw. Sci. Eng. (2018)
4. C. Troussas, A. Krouska, M. Virvou, Trends on sentiment analysis over social networks: pre-processing ramifications, stand-alone classifiers and ensemble averaging, in *Machine Learning Paradigms* (Cham, Springer, 2019), pp. 161–186
5. A. Krouska, C. Troussas, M. Virvou, Comparative evaluation of algorithms for sentiment analysis over social networking services. J. Univers. Comput. Sci. **23**(8), 755–768 (2017)
6. C. Troussas, A. Krouska, M. Virvou, Evaluation of ensemble-based sentiment classifiers for twitter data, in *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)* (2016), pp. 1–6
7. H. Thakkar, D. Patel, Approaches for sentiment analysis on twitter: a state-of-art study (2015). arXiv:1512.01043
8. Q.T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat, A. Rehman, Sentiment analysis using deep learning techniques: a review. Int. J. Adv. Comput. Sci. Appl. **8**(6), 424 (2017)
9. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space (2013). arXiv:1301.3781
10. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in *Advances in Neural Information Processing Systems* (2013), pp. 3111–3119
11. J. Pennington, R. Socher, C. Manning, Glove: global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014), pp. 1532–1543
12. P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information (2016). arXiv:1607.04606
13. M. Cliche, BB_twtr at SemEval-2017 Task 4: twitter sentiment analysis with CNNs and LSTMs (2017). arXiv:1704.06125
14. H. Nguyen, M.L. Nguyen, A deep neural architecture for sentence-level sentiment classification in Twitter social networking, in *International Conference of the Pacific Association for Computational Linguistics* (2017), pp. 15–27
15. A. Severyn, A. Moschitti, Twitter sentiment analysis with deep convolutional neural networks. in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2015), pp. 959–962

16. A. Cocos, A.G. Fiks, A.J. Masino, Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J. Am. Med. Inform. Assoc. **24**(4), 813–821 (2017)
17. T. Wu, S. Liu, J. Zhang, Y. Xiang, Twitter spam detection based on deep learning, in *Proceedings of the Australasian Computer Science Week Multiconference* (2017), p. 3
18. Z. Jianqiang, G. Xiaolin, Z. Xuejun, Deep convolution neural networks for Twitter sentiment analysis. IEEE Access **6**, 23253–23260 (2018)
19. J. Wehrmann, W. Becker, H.E. Cagnini, R.C. Barros, A character-based convolutional neural network for language-agnostic Twitter sentiment analysis, in *International Joint Conference on Neural Networks (IJCNN)* (2017), pp. 2384–2391
20. P. Vateekul, T. Koomsubha, A study of sentiment analysis using deep learning techniques on Thai Twitter data, in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (2016), pp. 1–6
21. H. Saif, M. Fernandez, Y. He, H. Alani, Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold, in *1st International Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013)* (2013)
22. D.A. Shamma, L. Kennedy, E.F. Churchill, Tweet the debates: understanding community annotation of uncollected sources, in *Proceedings of the First SIGMM Workshop on Social Media* (2009), pp. 3–10
23. M. Speriosu, N. Sudan, S. Upadhyay, J. Baldridge, Twitter polarity classification with label propagation over lexical links and the follower graph, in *Proceedings of the First Workshop on Unsupervised Learning in NLP* (2011), pp. 53–63
24. A. Krouska, C. Troussas, M. Virvou, The effect of preprocessing techniques on twitter sentiment analysis, in *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)* (2016), pp. 1–5
25. J. Zhang, C. Zong, Deep neural networks in machine translation: an overview. IEEE Intell. Syst. **30**(5), 16–25 (2015)