ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 4, April 2017**

# Ranking Based Reviews on Word Emotions

R.Rajani , P.Alekhya, Sk.Abbas

Head & Associate Professor, Dept. of MCA, Narayana Engineering College, Nellore, AP, India

Student, Dept. of MCA, Narayana Engineering College, Nellore, AP, India

Student, Dept. of MCA, Narayana Engineering College, Nellore, AP, India

**ABSTRACT**: We propose learning sentiment-specific word embeddings dubbed sentiment embeddings in this project. Existing word embedding learning algorithms typically only use the contexts of words but ignore the sentiment of texts. It is problematic for sentiment analysis because the words with similar contexts but opposite sentiment polarity, such as good and bad, are mapped to neighbouring word vectors. We address this issue by encoding sentiment information of texts (e.g. sentences and words) together with contexts of words in sentiment embeddings. By combining context and sentiment level evidences, the nearest neighbours in sentiment embedding space are semantically similar and it favours words with the same sentiment polarity. In order to learn sentiment embeddings effectively, We develop a number of neural networks with tailoring loss functions, and collect massive texts automatically with sentiment signals like emotions as the training data. Sentiment embeddings can be naturally used as word features for a variety of sentiment analysis tasks without feature engineering. We apply sentiment embeddings to word-level sentiment analysis, sentence level sentiment classification and building sentiment lexicons.

**KEYWORDS**: Sentiment - specific word embeddings, sentiment polarity, neural networks, sentiment classification.

## I.INTRODUCTION

Word representation attempts to represent aspects of word meanings. For example, the representation of “cell phone” may capture the facts that cell phones are electronic products, that they include battery and screen, that they can be used to chat with others, and so on. Word representation is a critical component of many natural language processing systems [4], [5] as word is usually the basic computational unit of texts.

A straight forward way is to represent each word as a one-hot vector, whose length is vocabulary size and only one dimension is 1, with all others being 0. However, one-hot word representation only encodes the indices of words in a vocabulary, but fails to capture rich relational structure of the lexicon. To solve this problem, many studies represent each word as a continuous, low-dimensional and real-valued vector, also known as word embeddings [6], [7], [8]. Existing embedding learning approaches are mostly on the basis of distributional hypothesis [9], which states that the representations of words are reflected by their contexts. As a result, words with similar grammatical usages and semantic meanings, such as “hotel” and “motel”, are mapped into neighbouring vectors in the embedding space. Since word embeddings capture semantic similarities between words, they have been leveraged as inputs or extra word features for a variety of natural language processing tasks, including machine translation [10], syntactic parsing [11], question answering [12], discourse parsing [13], etc all

## II. RELATED WORK

Word representation aims to represent aspects of word meaning. A straight-forward way is to encode a word $w_i$ as a one-hot vector, whose length is vocabulary size with 1 in the $w_i^{th}$ position and zeros everywhere else. However, such one-hot word representation only encodes the indices of words in a vocabulary, without capturing rich relational structure of the lexicon. One common approach to discover the similarities between words is to learn a clustering of words [25], [26]. Each word is associated with a discrete class, and words in the same class are similar in some respects. This leads to a one-hot representation over a smaller vocabulary size. Instead of characterizing the similarity with a discrete variable based on clustering results which corresponds to a soft or hard partition of the set of words, many researchers target at learning a continuous and real-valued vector for each word, also known as word

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

embeddings. Existing embedding learning algorithms are mostly based on the distributional hypothesis [9], which states that words in similar contexts have similar meanings. Many matrix factorization methods can be viewed as modeling word representations. For example, Latent Semantic Indexing (LSI) [27] can be regarded as learning a linear embedding with a reconstruction objective, which uses a matrix of “term-document” co-occurrence statistics, e.g. each row stands for a word or term and each column corresponds to an individual document in the corpus. Hyperspace Analogue to Language [28] utilizes a matrix of “term-term” co-occurrence statistics, where both rows and columns correspond to words and the entries stand for the number of times a given word occurs in the context of another word.

## III. SCOPE OF RESEARCH

In this paper, We propose learning sentiment-specific word embeddings dubbed **sentiment Embeddings** for sentiment analysis. We retain the effectiveness of word contexts and exploit sentiment of texts for learning more powerful continuous word representations. By capturing both context and sentiment level evidences, the nearest neighbours in the embedding space are not only semantically similar but also favour to have the same sentiment polarity, so that it is able to separate good and bad to opposite ends of the spectrum. In order to learn sentiment embeddings effectively, We develop a number of neural networks to capture sentiment of texts (e.g. sentences and words) as well as contexts of words with dedicated loss functions. We learn sentiment embeddings from tweets[1], leveraging positive and negative emoticons as pseudo sentiment labels of sentences without manual annotations. We obtain lexical level sentiment supervision from Urban Dictionary[2] based on a small list of sentiment seeds with minor manual annotation.
We evaluate the effectiveness of sentiment embeddings empirically by applying them to three sentiment analysis tasks. Word level sentiment analysis on benchmark sentiment lexicons [19], [20] can help us see whether sentiment embeddings are useful to discover similarities between sentiment words. Sentence level sentiment classification on tweets [21], [22] and reviews [23] help us understand whether sentiment embeddings are helpful in capturing discriminative features for predict the sentiment of text. Building sentiment lexicon [24] is useful for measuring the extent to which sentiment embeddings improve lexical level tasks that need to find similarities between words. Experimental results show that sentiment embeddings consistently outperform context-based word embeddings, and yields state-of-the-art performances on several benchmark datasets of these tasks.

## IV.PROPOSED SYSTEM

The proposed methods differ from context-based models in that we capture sentiment information of texts, which provides crucial evidences for capturing similarities between sentiment words. The hybrid models can be viewed as the “joint” version of ReEmbed by simultaneously encoding contexts of words and sentiment of sentences into word representation from scratch. Two hybrid models yields best performances as they capture not only contexts of words but also sentiment information of sentences.
The major contributions of the work presented in this paper are as follows :

- We propose learning sentiment embeddings that en-code sentiment of texts in continuous word representation.
- We develop a number of neural networks with tailoring loss functions to learn sentiment embeddings. We learn sentiment embeddings from tweets with positive and negative emoticons as distant-supervised corpora without any manual annotations.
- We verify the effectiveness of sentiment embeddings by applying them to three sentiment analysis tasks. Empirical experimental results show that sentiment embeddings outperform context-based embeddings on several benchmark datasets of these tasks.

## V. IMLEMENTATION

### 1. WORD LEVEL SENTIMENT ANALYSIS

we investigate whether sentiment embeddings are useful for discovering similarities between sentiment words in this section. we conduct experiments on word level sentiment analysis in two settings, namely querying neighbouring sentiment words in embedding space (Section 4.1) and word level sentiment classification (Section 4.2).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

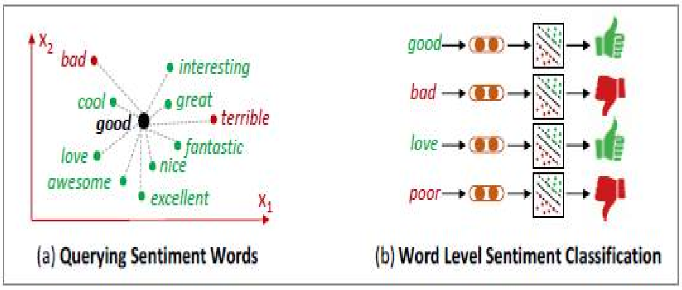
**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 4, April 2017**

**1.1 Querying Sentiment Words*:*** A better sentiment embedding should have the ability to map positive words into close vectors, to map negative words into close vectors, and to separate positive words and negative words apart. Accordingly, in the vector space of sentiment embedding, the neighbouring words of a positive word like "good" should be dominated by positive words like "cool", "awesome", "great", etc., and a negative word like "bad" should be surrounded by negative words like "terrible" and "nasty". Based this consideration, we query neighbouring sentiment words in existing sentiment lexicon to investigate whether sentiment embeddings are helpful in discovering similarities between sentiment words.




**1.2 Word Level Sentiment Classification:** We conduct word level sentiment classification to further investigate the effectiveness of sentiment embeddings in capturing similarities between sentiment words. Specifically, we conduct binary polarity classification on three sentiment lexicons (BL, MPQA and NRC) to infer whether a sentiment word expresses a positive or a negative meaning. The continuous word embeddings of a word are considered as its features for classification.

**2 .SENTENCE LEVEL ANALYSIS :**

In this part, we apply sentiment embedding as features to sentiment level sentiment classification. This helps us to investigate whether sentiment embedding is capable of capturing discriminative features for classifying the polar-ity labels (e.g. thumbs up or thumbs down) of text. We first present our strategy of using sentiment embedding as features for sentiment classification. We then describe experimental settings and empirical results.

We apply sentiment embeddings in a supervised learning framework for sentiment classification of sentences . Instead of using hand-crafting features, we use sentiment embeddings to compose the feature of a sentence. The sentiment classifier is built from sentences with manually annotated sentiment polarity .Specifically, we use a semantic composition based frame-work to get sentence representation. The basic idea is to compose sentence level features from sentiment embed-dings of words. This is based on the principal of compositionality , which states that the meaning of a longer expression (e.g. a sentence) is determined by the meaning of words it contains.

## III.BULDING SENTIMENT LEXICON

We apply sentiment embeddings to building sentiment lex-icon, which is useful for measuring the extent to which sentiment embeddings improve lexical level tasks that need to find similarities between words. We introduce a classification approach to build sentiment lexicon by regarding sentiment embeddings as word features, and then describe experimental settings and the results.

**3.1 A Classification Approach for Building Sentiment Lexicon:** We describe a classification approach for building large-scale sentiment lexicon, which is illustrated in Figure 8. We cast sentiment lexicon learning as a word-level classification task, which consists of two part: (1) an embedding learning algorithm to effectively learn the continuous representation of words, which are used as features for word-level sentiment classification, (2) a seed expansion algorithm that expands a small list of sentiment seeds to collect training data for building the word-level classifier. The learned sentiment embeddings are naturally regarded as continuous word features.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

*Website: www.ijircce.com*

**Vol. 5, Issue 4, April 2017**

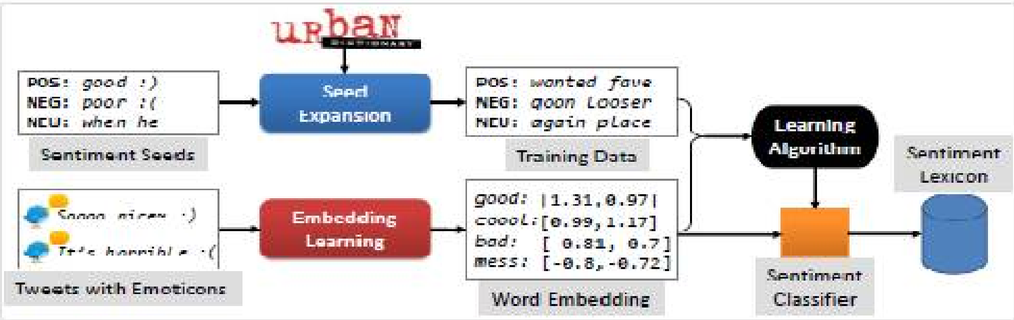


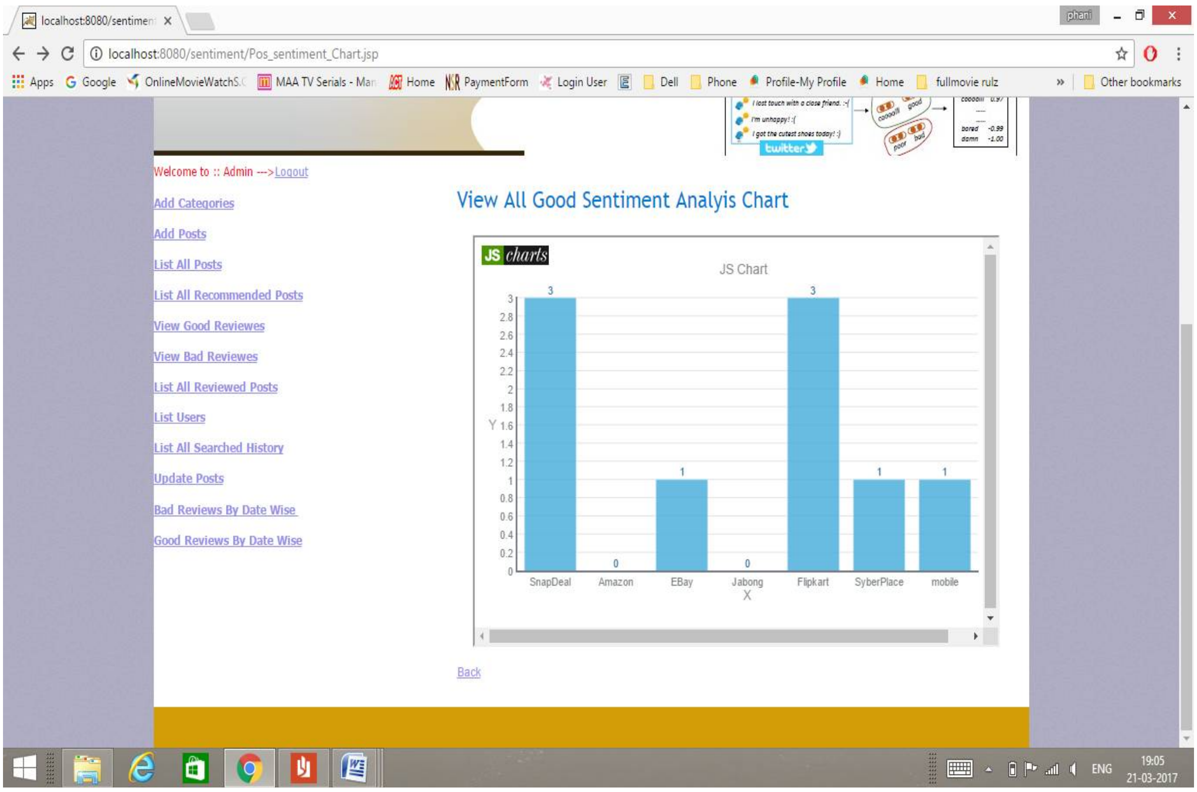

Fig. classification approach for building sentiment lexicon.

After obtaining the training data and feature representation of words, we build a word-level sentiment classifier with softmax, whose length is two for the positive v/s negative case:

$$y(w) = \text{softmax}( a. e_i + b)$$

where and b are the parameters of classifier, $e_i$ is the embedding of a word $w_i$, y(w) is the predicted sentiment distribution of $w_i$. We employ the classifier to predict the sentiment distribution of each word in the vocabulary of sentiment embeddings, and save the words as well as their sentiment probability in the positive (negative) lexicon if the positive (negative) probability is larger than 0:5.

## VI. EXPERIMENTAL RESULTS



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

***Website: www.ijircce.com***

**Vol. 5, Issue 4, April 2017**

**Fig:** In this above screen we can see that admin can see all positive comments which are given by the user in the way of view all good sentiment analysis chart.

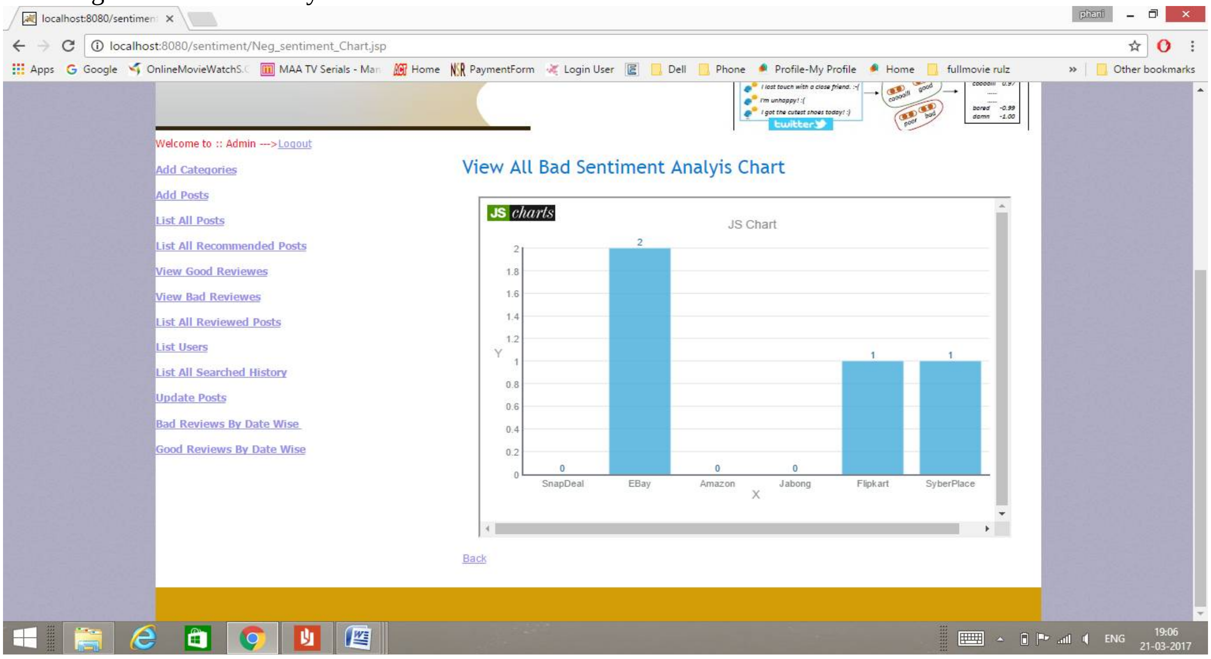



**Fig:** In this above screen we can see that admin can see all negative comments which are given by the user in the way of view all bad sentiment analysis chart.

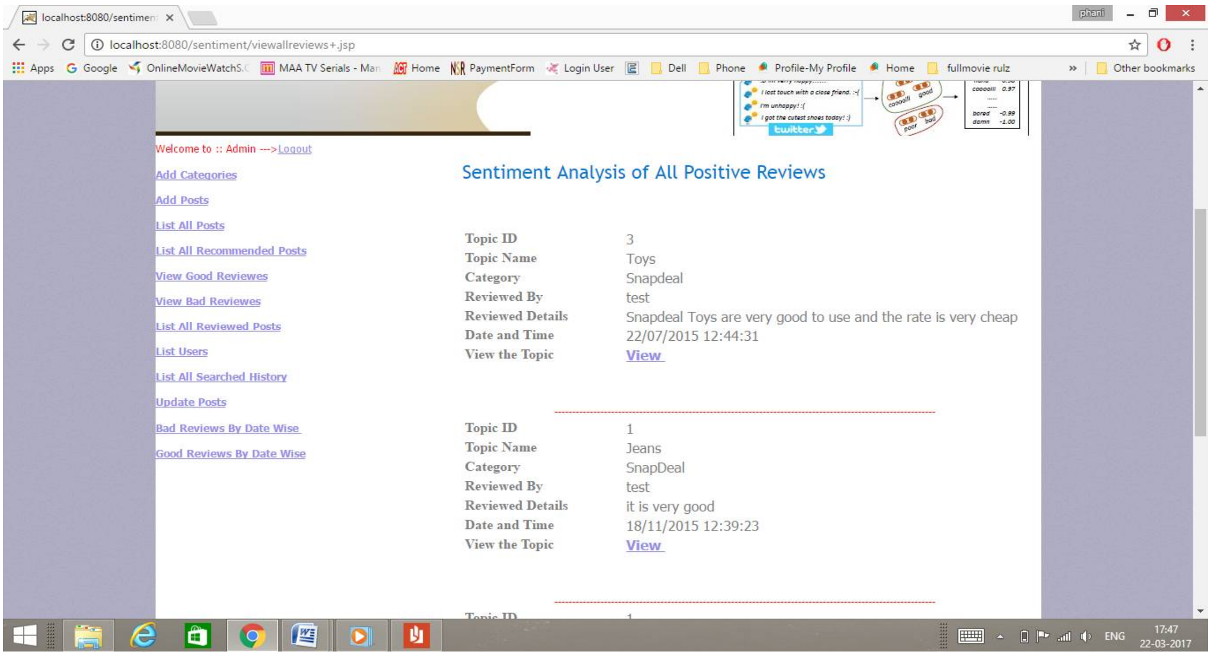



**Fig:** In this above screen we can see that admin can see all positive reviews which are given by the user in different categories and gives the information to other users about that product.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

***Website: www.ijircce.com***

**Vol. 5, Issue 4, April 2017**


**Fig:** In the above screen we can see that admin can see all negative reviews which are given by the user in different categories and gives the information to other users about that product.

## VII. CONCLUSION

We learn sentiment-specific word embeddings (named as sentiment embeddings) in this paper. Different from majority of exiting studies that only encode word contexts in word embeddings, we factor in sentiment of texts to facilitate the ability of word embeddings in capturing word similarities in terms of sentiment semantics. As a result, the words with similar contexts but opposite sentiment polarity labels like "good" and "bad" can be separated in the sentiment embedding space. We introduce several neural networks to effectively encode context and sentiment level informations simultaneously into word embeddings in a unified way. The effectiveness of sentiment embeddings are verified empirically on three sentiment analysis tasks. On word level sentiment analysis, we show that sentiment embeddings are useful for discovering similarities between sentiment words.

## REFERENCES

[1] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in Procedding of the 52th Annual Meeting of Association for Computational Linguistics., 2014, pp. 1555–1565.

[2] D. Tang, F. Wei, B. Qin, M. Zhou, and T. Liu, "Building large-scale twitter-specific sentiment lexicon : A representation learning approach," in Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, 2014, pp. 172–182.

[3] 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 2014), 2014, pp. 208–212.

[4] J. Pennington, R. Socher, and 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.

[5] N. Yang, S. Liu, M. Li, M. Zhou, and N. Yu, "Word alignment mod-eling with context dependent deep neural network." in Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013, pp. 166–175.

[6] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, "Parsing with compositional vector grammars," in Annual Meeting of the Association for Computational Linguistics, 2013.

[7] M. Iyyer, J. Boyd-Graber, L. Claudino, R. Socher, and H. Daume´ III, "A neural network for factoid question answering over para-graphs," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 633–644.

[8] J. Li, R. Li, and E. Hovy, "Recursive deep models for discourse parsing," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 2061–2069.

[9] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson, "Semeval-2013 task 2: Sentiment analysis in twitter," in Proceedings of the International Workshop on Semantic Evaluation, vol. 13, 2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

[10] A. Severyn and A. Moschitti, “Twitter sentiment analysis with deep convolutional neural networks,” in Proceedings of Special Interest Group on Information Retrieval (SIGIR). ACM, 2015, pp. 959–962.
[11] X. Hu, J. Tang, H. Gao, and H. Liu, “Unsupervised sentiment analysis with emotional signals,” in Proceedings of the International World Wide Web Conference, 2013, pp. 607–618.
[12] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in Proceedings of Language Resources and Evaluation Conference, vol. 2010, 2010.
[13] K.-L. Liu, W.-J. Li, and M. Guo, “Emoticon smoothed language models for twitter sentiment analysis,” in The Association for the Advancement of Artificial Intelligence, 2012.
[14] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, “Recursive deep models for semantic compositional-ity over a sentiment treebank,” in Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1631–1642.
[15] J. Li and E. Hovy, “A model of coherence based on distributed sentence representation,” in Proceedings of Empirical Methods in Natural Language Processing, 2014.
[16] J. Li, T. Luong, D. Jurafsky, and E. Hovy, “When are tree structures necessary for deep learning of representations?” in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 2304–2314.
[17] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word– emotion association lexicon,” Computational Intelligence, vol. 29, no. 3, pp. 436–465, 2013.

## BIOGRAPHY

Mrs. R. RAJANI is an Associate Professor and heading the department of MCA, Narayana Engineering College, Nellore, AP, India. She is pursuing her Ph.D from Sri Padmavathi Mahila University. She guided many projects for B.Tech and PG students. Her research interests include Datamining, Query Optimization, Computer Networks and Software Engineering etc.

Ms. P. Alekhya is a student pursuing MCA at Narayana Engineering College, Nellore, AP, India. During My final semester project work we prepared this paper for publishing it in an international journal.

Mr. Sk. Abbas is a student pursuing MCA at Narayana Engineering College, Nellore, AP, India. During My final semester project work we prepared this paper for publishing it in an international journal.