A Framework for Sentence Level Sentiment Analysis

Anjana P[1*], Syam Sankar[2]

[1]PG Scholar, Computer Science and Engineering, NSS College of Engineering, Kerala, India
[2]Assistant Professor, Computer Science and Engineering, NSS College of Engineering, Kerala, India

Abstract: ***Sentiment Analysis is one of the growing research fields in natural language processing. This paper proposes a framework for sentence level sentiment analysis it uses a joint segmentation and classification framework for sentence level sentiment analysis with recursive neural networking concept. The main issue of sentiment analysis is, in sentiment analysis a sentence { e.g. This movie is not bad } is consider as a negative polarity this is because of in sentiment analysis it consider like a sentence is split up in to words or word sequences. So we cannot predict the actual or correct polarity. To overcome this error here introduced a joint frame work for sentence level sentiment analysis. For training it uses the movie review dataset from IMDB.***

Keywords: Sentiment analysis, Classification, Recursive neural networking, Natural Language Processing

1. Introduction

Sentiment analysis is also called Opinion Mining (OM), sentiment analysis analyses peoples Opinion from text. It is based on the emotions of the people. It is an active research field n natural language processing. Natural language processing which processes the spoken language. Sentence Level sentiment analysis is a major part of sentiment analysis. Sentiment analysis has many types such as document level, aspect level, entity level etc.

The limitation of existing sentiment analysis is it split up sentence in to words it will reduces the efficiency of the result. In sentiment analysis it determines the sentiment polarity of a sentence based on its textural content. Mainly polarity considered as negative polarity and positive polarity. For a sentence e.g. that is bad, the overall polarity of this sentence is negative. Another e.g. that is good the polarity of this sentence is positive. The processing of sentiment analysis is mainly considered as lexicon based methods and corpus based methods. In lexicon based method [5]. **Lexicon based** approaches which are annotated with its sentiment polarity or sentiment strength. Linguistic rules such as intensifications and negations are usually ineffective to aggregate the sentiment polarity of sentences. **Corpus based** methods which considered sentiment classification as a text categorization problem. They mostly build sentiment classifier from sentences with annotated sentiment polarity.
Sentiment analysis is usually called Sentiment Analysis or Opinion Mining. Sentiment analysis involves classifying opinions in text into following categories like positive or negative or neutral. It is often named as subjectivity analysis, opinion mining [4]. Customers want to see the opinion of other about a product before they are purchasing it. In business organizations need to know what customers are telling about their services and products that an Organizations is providing to make future decisions.

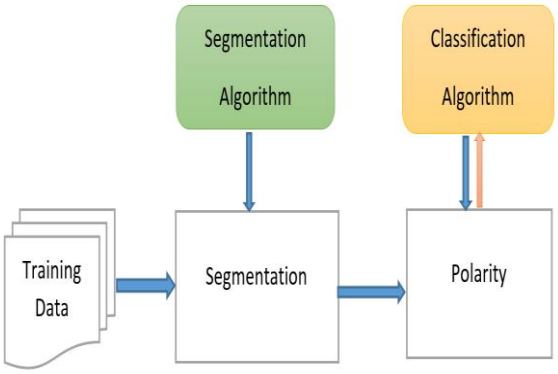

Figure 1. Pipelined Frame work

Sentiment analysis analyses the text. The fig.1 shows the pipelined method for sentiment analysis. It contains Segmentation and classification algorithm for processing of text. In segmentation stage they produce the segmentation result of a sentence using with bag-of-words or a separate text analyzer like standard syntactic chunker[6]. Second stage is the classification; it takes segmentation results as input and use a classification algorithm to build sentiment classifier. The classification algorithms are like supervised and unsupervised learning algorithms [1].

Recursive neural networking is a deep learning method. Which is an example of supervised learning method. In recursive neural networking, a recursive neural network is created by applying the same set of weights recursively over a distinguishable graph like structure, by traversing the structure in topological order.

Such networks are typically also trained by the reverse mode of automatic differentiation. They were introduced to learn distributed representations of structure, such as logical terms. A special case of recursive neural networks is the RNN itself whose format corresponds to a linear chain.

2. Related Works

Sentiment analysis or classification which classifies a sentence is positive or negative. There are three levels of

analysis tasks of Sentiment Analysis they are document level, sentence level and entity level. In document level sentiment analysis, it considers that a document is an opinion on an entity or aspect of it so this level called document-level sentiment classification. This idea is to classify whether the opinion document expresses a positive opinion or negative opinion sentiment. Examples like in product review, the system check whether the review expresses an overall positive or negative opinion about a particular product item. This task is commonly known as document-level sentiment classification [8]. In sentence level task at this level goes to the sentences and determines that the sentence expressed as a positive opinion, negative opinion, or neutral opinion. Neutral usually means unemotional opinion. This level of sentiment analysis is related to the subjectivity classification which processes sentences that express factual information from sentences that express subjective views and opinions. Aspect level act as a finer grained analysis. Aspect level was earlier called feature level. Instead of looking at language constructs, in aspect level sentiment classification it directly analyses the opinion itself. It is based on the idea that an opinion consists of a sentiment like positive or negative and a target opinion. It is mostly related to Feature-based Opinion Mining.

The existing method for sentiment analysis mainly two steam directions. Lexicon based approaches and corpus based approaches. Sentiment analysis is an active research field of natural language processing [9].

In lexicon based approaches it uses the existing sentiment lexicons of words and phrases, each of which is attached with the sentiment polarity or sentiment strength. Lexicon based method is an unsupervised learning method. The limitations of existing sentiment analysis is error propagation that is, the processing of sentiment analysis contains two stages. First one is the segmentation stage and second one is classification stage. In segmentation stage a sentence is segmented in to words here comes the tokenization concept. POS-tagging, stop words removal etc.

2.1 Machine learning approaches

Machine learning approaches can be divided in to two categories supervised and unsupervised machine learning techniques. The success of both is mainly based on the selection and extraction of the suitable set of features used to find out sentiments. In this method Natural Language Processing techniques play the most relevant position. Support vector machines (SVM), Naive Bayes, Maximum Entropy are some of the most common techniques used supervised learning approach[7].

2.2 Semantic Orientation Approach

The semantic orientation technic performs classification based on positive and negative sentiment words and phrases contained in processing of each text. It does not need any crucial training method to mine the data. There are two types of approaches are used in early sentiment analysis method using for the semantic orientation based approaches[16].

2.3 The corpus-based techniques

Corpus based techniques try to find out the occurrence patterns of words to determine their sentiments or polarity. General approach to build the sentiment dictionaries leverages both the language structure of and the corpus of text itself. In corpus based method builds the dictionary from the special words used in the texts from which sentiment is to be extracted and calculated a phrase's semantic orientation to be the mutual information between the phrase and the word "excellent" minus the mutual information between the phrase and the word "poor".

2.4 The dictionary-based techniques

Dictionary based techniques uses the synonyms, antonyms and hierarchies in WordNet or other lexicons with sentiment information to determine word sentiments [10].

2.5 Lexicon-based approaches

Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., it is a group of known sentiment terms, phrases and even idioms, developed for traditional method of communication such as the Opinion Finder lexicon but even more complex structures like Ontologies, or dictionaries measuring the semantic orientation of words or phrases can be used for this purpose.

2.6 Other Unsupervised Approaches

Bootstrapping is another approach. The idea is to use the output of an available initial classifier to create labeled data, to which a supervised learning algorithm may be applied. This method was used in conjunction with an initial high-precision classifier to learn extraction patterns for subjective expressions. Pang and Lee experiment with a different type of unsupervised approach. The problem they consider is to rank search results for review-seeking queries so that documents that contain evaluative text are placed ahead of those that do not. They propose a simple "blank slate" method based on the rarity of words within the search results that are retrieved (as opposed to within a training corpus).

3. Methodology

A typical kind of error is caused by the inconsistent sentiment polarity between a phrase and the words it contains, such as {“not bad, ” “bad”} and {“a great deal of, ” “great”}. The bag-of-words and syntactic chunkers are not effective enough to handle this polarity inconsistency phenomenon. For example, bag-of-words segmentation regards each word as a separate unit, which does not well capture the phrasal sentiment information like “not bad.” Syntactic chunkers typically aim to identify noun groups, verb groups or named entities from a sentence. However, many sentiment indicators are phrases constituted of adjectives, negations or idioms, which are divided by standard syntactic chunkers[12].

Here in fig.2 shows the proposed approach a new joint framework for sentence level sentiment analysis. It contains 2 phases prediction phase and training phase.

Besides, a better approach would be to utilize sentiment information as supervision to update the segment or, so that it will facilitate the performance of sentiment classification in turn. The drawback of existing sentiment classification algorithms typically split a sentence as a sequence of word, which could not successfully carry the inconsistent sentiment polarity between a phrase and the words.

Here a candidate generation model to produce candidates of each sentence; a segmentation ranking model to score the usefulness of a segmentation candidate for sentiment classification and a classification model for predicting the sentiment polarity of a segmentation. In the prediction process, the phrasal information of top-ranked segmentation candidates can be utilized as features to predict the sentence polarity.

Accordingly, the negation or polarity shifting expressions might be elegantly handled, as they are potentially grouped as basic computational units in segmentation results. In the training process, the segmentation ranking model is optimized with a marginal log-likelihood objective, which is designed for obtaining a better classification performance. This sentiment-specific segmentation ranking model can be effectively trained from sentences with only sentiment polarity labels, without using any syntactic or polarity annotations in segmentation level.

The proposed contains two phases prediction phase and training phase. In prediction process of this framework for sentence sentiment classification. The training process of this framework for sentiment classification. CG represents the candidate generation model, SC means the sentiment classification model and segmentation ranking model. The candidate generation model segmentation candidates are developed. These segmented candidates are in phrase level. So use a phrase level dictionary with beam search approach. Here use phrases identification by n-gram method. After obtaining the phrase dictionary apply beam search approach for each segmentation candidates. Another phase is segmentation ranking in this phase it assign score to segmentation candidate. Here use a log linear model to calculate segmentation score. In the training phase it optimize or updates the segmentation score using maximum likely hood optimization. Sentiment Classification phase build a classifier based on the recursive neural network method.

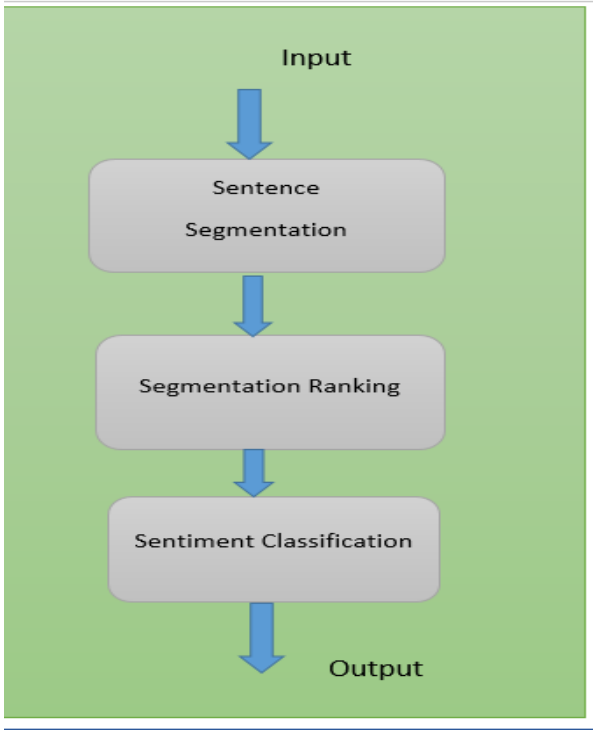

Figure 2. New Proposed approach

3.1 Candidate Generation Model

Candidate generation model is the first stage for this proposed frame work. In candidate generation model uses the beam search approach for the generation of segmentation candidates. Beam search approach, it will make the best candidate. Beam search approach it will segmented the sentence. It will make the possible segmentations with constraints on a phrase dictionary it induced from massive corpora. By using beam search approach it makes

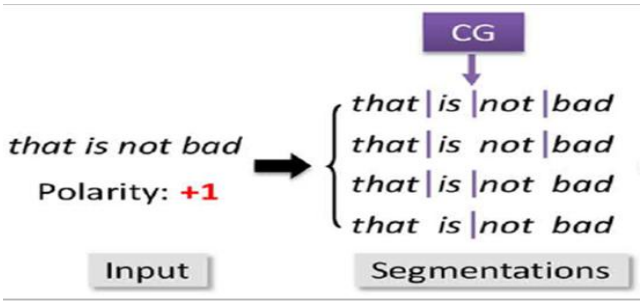

Figure 3 Candidate generation model

By implementing the beam search approach it makes basic segmentation units. It will is used for the segmentation ranking model. In this proposed method could not uses any syntactic analysis. Beam search will give best candidate. This candidate is used for the sentiment analysis.

3.2 Segmentation Ranking Model

Segmentation ranking is the next phase. In this phase it assign score to segmentation candidate. Here use a log linear model to calculate segmentation score. In the training phase it optimize or updates the segmentation score using maximum likely hood optimization. The segmentation

ranking model calculate the ranking with a log linear model the equation is given in equation (1). Segmentation ranking model is used for rank the sentence based on the score.

$$\phi_{ij} = exp(b + \sum_{k} sfe_{ijk} \cdot w_k) \qquad (1)$$

Based on the score it will rank the sentence the best scored sentence will be taken for the sentiment classification.

3.3 Sentiment Classification

Sentiment classification is the final stage of this proposed frame work. Here uses the recursive neural networking [2] concept for sentiment classification. In recursive neural networking it applies the same set of weights recursively over a structure, to produce a structured prediction over variable-length input by traversing a given structure in topological order. RNNs have been successful in learning sequence and tree structures in natural language processing, mainly phrase and sentence continuous representations based on word embedding. RNNs have first been introduced to learn distributed representations of structure, such as logical terms Fig.2 Represents the Recursive neural network architecture [11].

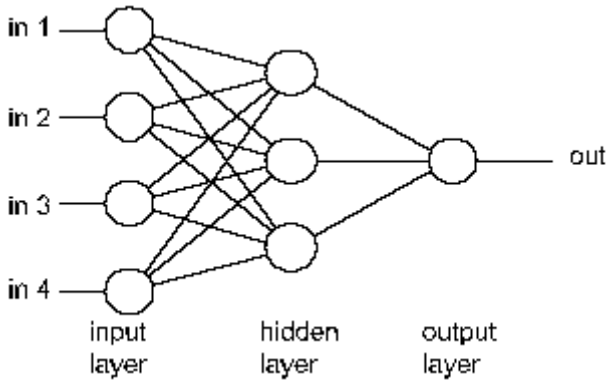

Figure 4 Artificial Neural Network

Feature extraction places an important role in natural language processing. The Fig.5 shows the Sentiment classification model for the proposed approach.

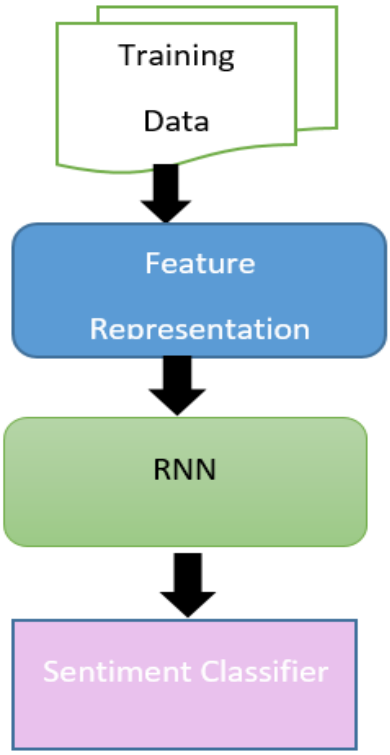

Figure 5. Sentiment classification

4. Conclusions

A sentence level sentiment analysis that analyses the overall polarity of a sentence positive like positive or negative polarity. Here uses the joint segmentation and classification frame work as a base method. This method uses the candidate generation, segmentation ranking and classification here candidate generation uses the beam search algorithm for best candidate generation. Uses marginal likely hood estimation for segmentation ranking and recursive neural network model for sentiment classification.

References

[1] D. Tang, F. Wei, B. Qin, L. Dong, T. Liu, and M. Zhou, “A joint segmentation and classification framework for sentiment analysis”, in Proc. EMNLP, pp. 477-487, 2014.
[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality”, Proc. Conf. Neural Inf. Process. Syst, 2013.
[3] G. Frege, “On sense and reference”, vol. 1892, pp. 563-584, 1997.
[4] B. Pang and L. Lee, “Opinion mining and sentiment analysis”, Foundations Trends Inf. Retrieval., vol. 2, no. 12, pp. 1-135, 2008.
[5] C. Havasi, E. Cambria, B. Schuller, B. Liu, and H. Wang, “Knowledge based approaches to concept level sentiment analysis”, IEEE Intell. Syst., vol. 28, no. 2, pp.0012-14, 2013.

[6] P. D. Turney, “Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews”, in Proc. ACL, pp. 417-424, 2002.
[7] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques”, in Proc. EMNLP, pp. 79-86, 2002.
[8] X. Ding, B. Liu, and P. S. Yu, ”A holistic lexicon-based approach to opinion mining”, in Proc. Int. Conf. Web Search & Datamining, pp.231-240.
[9] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision”, Stanford University, pp. 1-12, 2009.
[10] J. Mitchell and M. Lapata, “Composition in distributional models of semantics”, Cognitive Sci., vol. 34, no.8, pp.1388-1429, 2010.
[11] Y. Kim, “Convolutional neural networks for sentence classification”, 2014.
[12] J. Li, Feature weight tuning for recursive neural networks, Arxiv Preprint, pp.1412-3714, 2014.
[13] Vrushali, K. Bongirwar, “A Survey on Sentence Level Sentiment Analysis”, International Journal of Computer Science Trends and Technology (IJCST), Vol. 3, Issue 3, pp.78-95, 2015.
[14] S. Wang and C. D. Manning, “Baselines and bigrams: Simple, good sentiment and topic classification”, in Proc. ACL, pp. 90-94, 2012.
[15] H. D. Kim and C. Zhai, “Generating comparative summaries of contradictory opinions in text”, in Proc. CIKM'09, 2009.
[16] Tetsuya Nasukawa, and Jeonghee Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", K-CAP'03, pp. 70-77, October, 2003.
[17] M. Hu and B. Liu, "Mining and summarizing customer reviews, " in Proc. SIGKDD, pp. 168-177, 2004.
[18] Salvetti F., Lewis S., Reichenbach C. Automatic Opinion Polarity Classification of Movie Reviews. Colorado Research in Linguistics. Volume 17, Issue 1. Boulder: University of Colorado.2004.
[19] Pak, A. Paroubek, P. “Twitter as a corpus for sentiment analysis and opinion mining”, Proceedings of LREC 2010, Retrieved October 5, 2010.