

A Survey on Review Analysis Using Deep Learning Techniques

Aswathi Sajeevan

(Information Technology, Rajagiri School of Engineering and Technology, India)

Abstract: Sentiment Analysis is the branch of Text Mining which classify the emotions or opinions of a people based on their reviews on social media. Sentiment analysis is done with the help of Dictionary based methods and machine learning based methods. Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) are traditional methods for using sentiment analysis. Deep learning is a novel method which provides good definition for sentiment analysis with the help of various deep learning methods likes Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM).

Keywords: CNN, Deeplearning, LSTM, RNN

I. INTRODUCTION

The main idea behind the sentiment analysis is to predict the emotions or opinions of the people towards a particular topic from a structured, semi-structured or unstructured textual data. The text data contains human higher level emotional characteristics. With the rapid development of data people around world have a greater and greater passion for expressing their emotions and opinions on internet. Most of the industries and organizations to make a better decision based on their customer reviews on different social websites. The fast advancement of the text data, Sentiment analysis becomes important research direction in Natural language processing (NLP). Sentiment analysis plays a significant job in our everyday decision making process. In early days, before buying a product or service people will seek opinions from friends, relatives, neighbours etc. But in the recent days, people will get opinions from different websites such as twitter, face book in such situations sentiment analysis plays a crucial role.

Sentiment analysis is done with the assistance of various machine learning based approaches or dictionary based techniques. Sentiment classification can be performed in 3 phases named as Document level, Sentence level and Feature level. In Document level, find sentiment polarity for the whole review but extracting the sentiment for whole document/blog will not be efficient. So further moved to sentence level or feature level classification, which can express a sentiment polarity for each sentence of a review and even for each word.

II. SENTIMENT ANALYSIS METHODS

Sentiment analysis process starting with collecting reviews or data set. After collecting reviews next is the sentiment identification that means identify opinionative words or phrases. Important part of the sentiment analysis is feature selection, most commonly used features are term presence and their frequency, Negations, Part of speech information, assessment words and phrases. Finally sentiment classification process that is based on sentiment polarity. It is either 0 or 1, zero means that particular review belongs to negative class otherwise positive class. Sentiment analysis strategy can be divided in to dictionary based and machine learning based.

Dictionary based methods are two types one is wordNet and another is SentiWordNet. WordNet is a lexical database for the English language that combine English words in to set of synonyms called synset. It is differentiate between nouns, verbs, adjectives and adverbs. Next is the SentiWordNet is an extension of WordNet. SentiWordNet set sentiment scores for each synset of WordNet. Sentiment scores are either positive or negative. Object score can be calculated with the help of a positive and negative score. Summation of positive score and negative score subtracted from 1 we get object score. Dictionary based methods are executed through following steps [10] tokenization, speech tagging, word sense disambiguation (WSD), Sentiwordnet interpretation, sentiment orientation and tweet classification. Major disadvantage of dictionary based approach is inability to determine the opinion words with domain and context.

Machine learning based methods are classified in to two, supervised learning and unsupervised learning methods. Supervised learning use labelled training documents for training the model. The training data contains a pool of training examples. Each example in supervised learning comprises of an input object and output value. Unsupervised learning methods don't use the labelled documents. Labelled data means data consist with natural language or human artefacts like photos, news articles and audio recordings are unlabelled data while each set of unlabelled data with meaningful tag is called labelled data. For example photo is unlabelled data that photo contain apple is a labelled data. They create use of the document's statistical properties, for example, word co-occurrence, Natural Language Processing, existing dictionaries with passionate words.

III. MACHINE LEARNING METHODS

Different machine learning techniques have been implemented to categorize the sentiments. Machine learning techniques like Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved good success in review analysis. Machine learning begins with collecting training dataset. Important part of the machine learning method is feature selection. They describe how documents are denoted. Most commonly used features are term presence and their frequency, Negations, Part of speech information, Opinion words and phrases [5].

Support vector machine is a machine learning method used for classifying the text data. SVM can work with both linear and non-linear separable data. SVM perform pattern extraction and data classification operation with the help of a hyper plane. Support vector machine define decision boundaries based on decision planes. Decision planes are make separation between two classes. The main problem is that, it can't deal with the huge number of input data and training phase of SVM takes much more time than actual classification phase.

Naive Bayes is a probabilistic classifier method based on Bayes theorem. This method combines the probability of numerous independent events into one number. The conditional probability that an event A given B is determined by Bayes rule $P(A/B) = P(A) P(B|A) / P(B)$. High scalable nature and incremental learning is the strong point of Naive Bayes Method. A main issue deal with NB is, it is most depended on shape of data distribution and another problem due to data scarcity. Maximum Entropy (ME) is another probabilistic classifier method. Maximum entropy can be calculated with the help of Multivariate Information theory. ME find group of features from input text then joining them and find sum as its exponent.

Various machine learning based methods are used for sentiment classifications but their associated with some disadvantages. For example commonly used machine learning methods are Support vector machine (SVM), Maximum entropy (ME) and Naive bayes (NB). They each have advantages and disadvantages. SVM and ME are complex models and take longer training time. Naive bayes approach fails to take the order of the words from input text and hence leads to a less accurate model. Another modelling technique used is n-grams, but this has a fixed range and cannot be learn for long term dependencies. These approaches also often fail to classify sentences with a negation. For example, 'I like this novel' and 'I don't like this novel'. Here don't is act as a negation and get a negative label but these models are give it a positive label. These problems are overcome with the help of deep learning methods.

IV. DEEP LEARNING METHODS

Machine learning methods automatically detect the available pattern from a given set of text data. Deep learning is an addition of machine learning method. Deep learning is a powerful tool for finding emotions and opinions from huge text data. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) are the deep learning methods used for sentiment analysis. An Artificial Neural Network (ANN) is an information processing model that is simulated by biological nervous systems, that means contain lots of neurons and process information. The key element of this paradigm is large number of information processing units called neurons. An ANN is designed for a particular application, for example, hidden pattern recognition or data classification, through a learning procedure.

Artificial neural network consist of 3 layers named as input layer, hidden layer and output layer. Input layer, which links to the input vector and output layer, which corresponds to the output vector. Intermediate layer is called hidden layer, which perform some computation with the help of some activation functions. Basic computational element of neural network is called activation function [2]. Commonly used activation functions are sigmoid function, hyperbolic tangent function (tanh), or rectified linear function (ReLU). A line between two layer is denotes a connection for the flow of information. Each connection is associated with a value is called weight, that controls the signal between two neurons.

4.1 Convolutional Neural Network (CNN)

CNN is another type of deep neural network. Architecture of CNN consists of one input layer, multiple hidden layers which are convolutional layers, pooling layers, fully connected layers, and one output layer. CNN contain some Convolutional layers that apply a convolution operation to the input, and the result is passing to the next layer. A convolutional layer consist of various filters that does convolutional activity. The filter is really a variety of numbers (called weights or parameters).

Machine couldn't have ability to read human language or natural language such situation, the input text data represented as mathematically with the help of a word2vec. The output of the word2vec is served as a input to the CNN. Initially filters scan the input vector. Filter actually array of numbers. If the filter is sliding over the input vector, then multiplying its weight value with original value of the text. After filter finishes scanning over the input vector. We can get number of activation maps or feature maps. Reduce these feature maps with the help of subsampling or pooling layer. Subsampling layer is used to reduce the spatial size of the

representation thus to reduce the computational complexity of the network. Commonly used subsampling operation is maxpooling. Finally the output is goes to the fully connected layer of neural network for classification.

4.2 Recursive Neural Network (RNN)

RNN is type of neural network comprise of three modules, which are input layer, hidden layer and output layer. RNN can utilize its inside "memory" to process a succession of inputs. It is able to learn more information in long sequence compared with feed forward network. The pattern in sentences can be recorded in hidden layer and the model needs less parameters. The hidden layer successfully reserves all information in previous words, which improves the performance of identifying the sequential relationships between words. RNN contain loops to allow information to be persistent.

Main limitation of RNN is the vanishing gradient and exploding problems. Vanishing gradient problem means vanishing gradients where gradients tend to slowly disappear as we back propagate across multiple unfolding of the RNN layer. Due to the vanishing gradient problem the training of RNN difficult because it cannot process very long sequences if using tanh as its activation function and it is very unstable if using ReLU as its activation function. Another disadvantage of RNNs is not able to keep track of long term dependencies. LSTM can be used to overcome these issues.

4.3 Long Short-Term Memory (LSTM)

LSTM is a unique sort of RNN, which is equipped for adapting long term conditions. The main part of an LSTM model is a memory cell. LSTM deals with two types of memory Short term memory and Long term memory. This model allows remembering of information over a much longer period. The LSTM model consists of memory cells to maintain information over long periods of time with the help of gating units. The gating units are deciding what information to store and when to apply that information. Each memory cell divided in to 3 gates named as input gate, forget gate and output gate.

At time step t , LSTM first chooses what data to dump from the cell state. This choice is made by forget gate. The data that is never again required for the LSTM to comprehend things or the data that is of less significance is removed. $h(t)$ is the hidden state at time step t and $h(t-1)$ is the hidden state value at time step $t-1$. $x(t)$ is the current input value at time step t . LSTM chooses what new data to store in the cell state. This procedure for including some new data done via input gate. This contains two steps update and creates. First, a sigmoid function/layer, called the "input gate" as decides which values LSTM will update. Next, a tanh function creates a vector of new candidate values, which will be added to the cell state. Multiplying the value of the sigmoid function to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation. Then LSTM runs a sigmoid layer, which decides which parts of the cell state to output, called output gate.

V. RELATED DEEP LEARNING METHODS FOR SENTIMENT CLASSIFICATION

Deep learning is one of the efficient methods for implementing sentiment analysis, which have been expanded recently. These models include CNN (convolutional neural networks), RNN (recursive neural network) and LSTM (Long Short Term Memory Network). This section describes the efforts of different researchers toward implementing deep learning models for performing the sentiment analysis.

The vast majority of studies have utilized SVM, Naïve bayes and other traditional methods, which dependent on manual labelling of a lot of enthusiastic attributes, however paid a high cost. To enhance this circumstance, Liu Yanmei et al [14] additionally examined deep learning and Micro-blog sentiment analysis, and proposed deep learning CNN with SVM classifier. This Framework avoids problems due to the explicit feature extraction but many difficulties are faced in emotional analysis using deep learning. Author doesn't find good solution for this. This paper demonstrates that, this framework is good to enhance the accuracy of sentiment analysis, however there are as yet numerous issues in the study of emotional analysis Using deep learning, the following stage will be to additionally consider the deep learning techniques for finding a more reasonable sentiment analysis.

Yuzhou et al[13] described a CNN based Diversified Restrict Boltz- Mann Machine (RBM) technique to model the sequence level latent topics in the sentences for emotion analysis. Diversified RBM to find the latent topic .These topics in the sentence is used as the feature of sentiment classification and diminish the logical ambiguities but training process is more complex.

To overcome the difficulties of representing input text in to word vector, Zhang Yangsen et al[12] proposed CBOW model with RNN. It is better than Naive Bayes method, and can learn sentence semantic relation automatically and also effectively recognizes negative sentences and twofold negative sentences, and can work steadily as the length of sentence increments.

Dan Li et al [11] introduce a novel method LSTM to overcome vanishing gradient problem associated with RNN.LSTM, which successfully capture long term dependencies of sentences and gives better results than traditional RNN.Igor Santos et al [10] proposed a CNN variant named word2vec-CNN that employs input text as a word vector and serve this word vector as a input of the convolutional layer. This paper offers the usage of recently released Facebook fast text word embedding as illustration of word to perform the assignment of sentiment analysis. The CNNs have a few different hyper parameters and to tuning them is an extremely tedious errand. In this work, author kept them steady while changed the word embedding because they are needed to examine their influence on the execution of the CNN.

Ali MertErtugrulet [9] al use a bidirectional LSTM to perform movie genre classification from plot summaries of movies. This framework Works also in sentence level classification and Provide better prediction of context and eliminate ambiguity but Applied only for small data sets. HaoLidong et al [8] proposed a Multimixed CNN with cyclic matrix filling and random matrix filling framework. MMCNN to take care of the sparse problem of the word embedding matrix and also acts as a good feature learner but this applied only for small data sets.

Abdalraouf Hassan et al [6] introduce a combination of CNN and LSTM methods. This work use the advantages of both CNN and LSTM.CNN acts as a good feature learner but CNN doesn't capture long term dependencies .This problem overcome with the help of LSTM.LSTM layer is stacked above of convolutional layer .Output of the convolutional layer serve as a input to the LSTM.

VI. DEEP LEARNING METHODS USED FOR SENTIMENT CLASSIFICATION

Sr. No	Author	Year of Publication	Methods	Advantages	Disadvantages
1	Abdalraouf Hassan et al	2018	Hybrid CNN and LSTM	Meaningful feature extraction and capture long term dependency	Complex architecture
2	Jin Wang et al	2018	Hybrid CNN and SVM	Classification problem of CNN can be overcome with SVM	Only for small data sets
3	HaoLidong et al	2018	Multimixed CNN with cyclic matrix filling and random matrix filling	MMCNN is advantage from the feature combination which is constantly overlooked in different models. To take care of the sparse problem of the word embedding matrix.	Provide good accuracy but time consuming.
4	Ali MertErtugrul et al	2018	Bidirectional LSTM	Works also in sentence level classification and Provide better prediction of context and eliminate ambiguity	Applied only for small data sets.
5	Igor Santos et al	2017	Word2Vec and CNN	CNN can extract meaningful feature from input samples effectively	Classification ability of fully connected classification layer is weak for nonlinear separable data.
6	Dan Li et al	2016	LSTM	Capture Long term dependencies	Feature extraction is not possible in the case of large data set
7	Zhang Yangsen et al	2016	CBOW with RNN	Achieves a better accuracy compared to Naive Bayes model.	Vanishing gradient problem
	YU Zhou et al	2016	CNN based	Diversified RBM to find	Training process

8			Diversified Restrict Boltzmann Machine (RBM) method	the latent topic .These topics in the sentence is used as the feature of sentiment classification and diminish the logical ambiguities.	is more complex
9	Liu Yanmei et al	2015	deep learning CNN with SVM classifier	Avoid problems due to the explicit feature extraction	This framework not works properly. Many difficulties are faced in emotional analysis Using deep learning. Author doesn't find good solution for this.

Table 1: Deep learning methods used for sentiment classification

VII. CONCLUSION

In day by day the velocity and volume of the text data increment rapidly. Sentiment analysis plays an important roles in our daily decision making process. Sentiment analysis done with dictionary based methods and machine learning based method. In recent years deep learning methods provide better accuracy for sentiment classifications compared with traditional methods. In this paper, the deep learning methods in sentiment analysis in recent years are summarized such as CNN, RNN, LSTM, Bidirectional LSTM and also summarize traditional methods used for sentiment analysis(such as WordNet,Senti Word Net,SVM,NB and ME.

REFERENCES

- [1] Xiaobo Zhang, Qingsong Yu." Hotel Reviews Sentiment Analysis Based on Word Vector Clustering". 2017 2nd IEEE International Conference on Computational Intelligence and Applications.
- [2] A.M.Abirami,.V.Gayathri." A Survey On Sentiment Analysis Methods And Approach". 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).
- [3] Zhongkai Hu, Jianqing Hu, Weifeng Ding andXiaolinZheng "Review Sentiment Analysis Based on Deep Learning" 2015 IEEE 12th International Conference on e-Business Engineering
- [4] S. M. Vohra, J. B. Teraiya "A Comparative Study Of Sentiment Analysis Techniques" Journal Of Information, Knowledge And Research In Computer Engineering ,2016
- [5] YayaHeryadi ,Harco Leslie Hendric Spits Warnars." Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM". 2018 Conference on Information Communications Technology and Society (ICTAS).
- [6] Abdalraouf Hassan, AusifMahmood, "Deep learning for sentence classification",Systems, Applications and Technology Conference (LISAT), 2017 IEEE .
- [7] Yuling Chen, Zhi Zhang "Research on text sentiment analysis based on CNNs and SVM". 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).
- [8] Lidong, Hao, and Zhao Hui. "A new short text sentimental classification method based on multi-mixed convolutional neural network." 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, 2018.
- [9] Ali MertErtugrul and Pinar Karago "Movie Genre Classification from Plot Summaries using Bidirectional LSTM" 12th IEEE International Conference on Semantic Computing 2018
- [10] Igor Santos, Nadia Nedjah and Luiza de MacedoMourelle "Sentiment Analysis using Convolutional Neural Network with fastText Embeddings"2nd IEEE International Conference on Computational Intelligence and Applications 2017
- [11] Dan Li, Jiang Qian "Text Sentiment Analysis Based on LongShort-Term Memory" IEEE International Conference on Computer Communication and the Internet 2016
- [12] Zhang Yangsen, Jiang Yuru and Tong Yixuan "Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network" Chinese Journal of Electronics Vol.25, No.4, July 2016

- [13] Zhou, Yu, RuifengXu, and Lin Gui. "A sequence level latent topic modeling method for sentiment analysis via CNN based diversified restrict boltzmann machine." *Machine Learning and Cybernetics (ICMLC), 2016 International Conference on*. Vol. 1. IEEE, 2016
- [14] Liu Yanmei , Chen Yuda” *Research on Chinese Micro-blog Sentiment Analysis Based on Deep Learning” 8th International Symposium on Computational Intelligence and Design , IEEE 2015*