

STUDY OF DEEP LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS AND ASPECT EXTRACTION OF RESTAURANT REVIEWS

Abila Saravanabhavan, Akshara Saji, Balakirthika Esaikkiappan, Anushka Agashe
Final Year Bachelor of Engineering
Department of Computer Engineering
S.I.E.S. Graduate School of Technology,
University of Mumbai, Maharashtra ,India

Abstract: Machine learning has certainly changed how data is perceived and utilized. With this progress the fine dining experience has leapt into a new phase. With social media, people have started being more vocal about their experiences about dining at restaurants by writing reviews. Sentiment analysis of the reviews determines the polarity of the population towards various restaurants. Along with it, aspect extraction deduces the relevant features about the restaurants as stated in the reviews. In this project we exploit deep learning algorithms to extract the general sentiment and classification of all specific features of restaurants in the reviews. We propose a Recurrent neural network (RNN) - Bidirectional LSTM model to do classification and aspect extraction of reviews by Convolutional neural network (CNN) algorithm. Also we have conducted a comparative analysis of RNN and CNN algorithms for sentiment analysis.

Keywords : Aspect based sentiment analysis (ABSA), Bidirectional Long Short Term Memory (Bi - LSTM), Convolutional neural network (CNN), Deep learning (DL), Natural Language Processing (NLP), Recurrent neural network (RNN), Sentiment analysis (SA).

I. INTRODUCTION

Sentiment is a view or an opinion of a person with regards to a particular product or service .Sentiment Analysis is the process of analyzing the text which either in textual or audio format to extract various sentimental/emotional values from the text for understanding people's perception towards a particular subject and to build a opinion based on the results of the analysis. In other words, it is an automated system which uses NLP (Natural Language Processing) techniques. Aspect Extraction (ABSA) is the extraction of particular aspect of the text content the users have commented describing about a particular subject. The SA Algorithm predicts whether the given text content has a positive or negative polarity. The project processes the reviews regarding a particular restaurant from the users ,extracting the sentimental value from the reviews (i.e. positive , negative) which benefits the user to make a conclusion and the restaurant owners to plan the strategy for business development .The reviews will be further analyzed on aspects/features for judging a restaurant based on user reviews .

II. LITERATURE SURVEY

The different types of techniques available for Sentiment Classification is based on Machine Learning are Supervised, Unsupervised and Semi-Supervised learning techniques [8]. Unsupervised learning techniques need a large amount of data for training which is a major drawback of unsupervised learning techniques. Supervised learning techniques is associated with drawbacks such as its sensitivity to quality and quantity of training data and its failure in case of biased training data.[8]. Though Learning Automata based Sentiment analysis can be theoretically explained, they have no proper practical implementation which leads us to the new technique i.e. Deep Learning using RNN (Recurrent Neural Network) [1]. The proposed paper[5] puts forth a comparative study of all the different Deep Learning techniques such as CNN(Convolution Neural Network) , RNN(Recursive Neural Network), DNN(Deep Neural Network) , RNN(Recurrent Neural Network),DBN(Deep Belief Network) and HNN(Hybrid Neural Network) leading to a conclusion that Deep Learning techniques are better and more accurate and efficient in Sentiment Analysis[5]. [4] A more improvised version of Sentimental Analysis algorithm i.e. RNN (recurrent Neural Network) along with LSTM (Long Short Term Memory) Network is explained in the proposed paper. The complete model of RNN with LSTM is put forth [4]. Joint segmentation and classification of opinions in text can be performed using RNN along with bi-directional LSTM. Two cost functions are used in the proposed model, one for segmentation and other for classification. These two cost functions are used to train the model to optimize the network. This method gives a perfect and accurate result in both cases (with the knowledge of full stops and commas and without the knowledge of full stops and commas)[4]. [9]Conventional SA of text focuses on the opinion of the entire text or the sentence. In the case of reviews, it has been observed that customers often talk about multiple aspects of an entity and express an opinion on each aspect separately rather than expressing the opinion towards the entity as a whole. ABSA has emerged to tackle this issue.[10]CNN model to predicts the specific category sentiment as an adjective may be a positive sentiment indicator for one category and negative for another.

III. METHODOLOGY

A. Pre-Processing

- 1. Data collection:** Data (restaurant reviews) is in the raw form. Approximately 80000 restaurant reviews, taken from kaggle.com are used each having a sentiment (positive or negative) and aspects (food_quality, restaurant_general, service_general) which are used for training and testing.
- 2. Data processing:** Data processing involves tokenization and normalisation.
Tokenization- Tokenization means isolating the reviews into individual words called tokens. Tokens can be split using punctuation or white spaces.
Review - I love food
Tokenized review - "I" "love" "food"
Normalization- The next step in data processing is normalization i.e. conversion of reviews into lowercase. Reviews are normalized by converting it to lowercase which makes its comparison with a dictionary easier.
- 3. Bag of Words:** A matrix of occurrence of all words in the vocabulary is made.
- 4. Noun Chunking:** The nouns in the review are extracted using parsers in spaCy.

B. Sentiment Analysis using RNN

Recurrent Neural network (RNN) -It is a class of artificial neural network where connections between nodes form a directed graph through sequence. They are neural networks with loops in them, allowing information to sustain.

Bidirectional Long Short Term Memory Networks (Bi- LSTM) -It trains two LSTMs on the input sequence. The first on the input sequence as it is and the second is on the reversed copy of the input sequence. This provides an additional context to the network and result in faster and fuller learning on the problem. Using these algorithms the restaurant reviews are classified as positive or negative.

C. Sentiment Analysis using CNN

CNN is a class of deep neural network which is used to minimize the data preprocessing. A convolutional neural network consists of an input layer, an output layer and hidden layers. The hidden layers of a CNN typically consist of convolutional layers, RELU layer i.e. activation function, pooling layers, fully connected layers and normalization layers.

D. Aspect based Sentiment Analysis

A typical sentimental analysis predicts the polarity of the whole sentence. But in aspect based the features/ aspects are extracted and the polarity of those features are predicted. The aspects considered in restaurant reviews dataset are food_quality, restaurant_general, service_general.

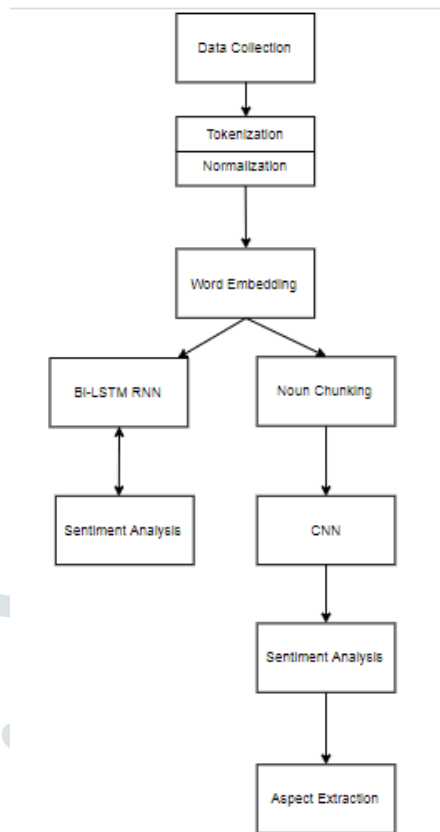


Fig.1 Methodology

IV. EXPERIMENTAL EVALUATION

The proposed system has been tested using Restaurant Reviews which include reviews classified into positive and negative sentiments. Accuracy of RNN is compared with CNN classifier for sentiment analysis.

Table 1: Accuracy comparison

| Classification Model | Accuracy |
|----------------------|----------|
| RNN | 75.90% |
| CNN | 84.32% |

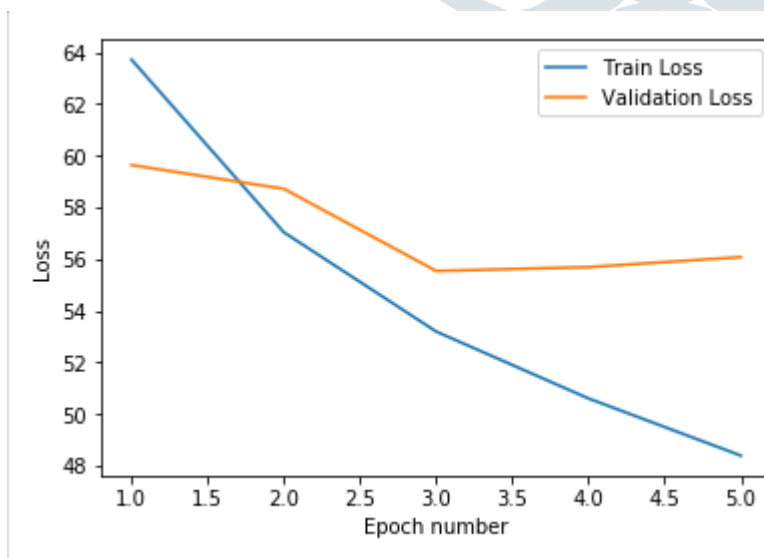


Fig.2 RNN-Train loss and accuracy against number of epochs.

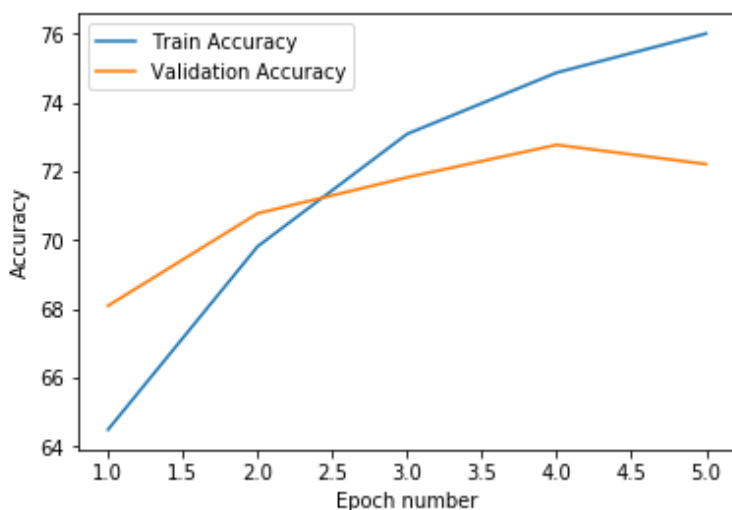


Fig.3 CNN-Train loss and accuracy against number of epochs.

From both the graphs it is clear that accuracy of CNN is more than the accuracy of RNN SA. Also as the number of epochs increases, accuracy of the algorithms increases in both the cases. The evaluation of both the models has led the understanding that CNN is slightly better than RNN.

V. CONCLUSION

In this project we have proposed RNN Bi-LSTM for SA and CNN for SA and ABSA on Restaurant reviews. The comparative study shows CNN has a better accuracy in RNN. Many further research can be done to increase the accuracy considering datasets, epochs, batch size, hidden layers.

VI. ACKNOWLEDGMENT

We would like to show our sincere gratitude to Dr. Vikram Patil, our Principal for motivating us to think big and come up with an idea which will help people make better choices and lead a satisfied life. Our deep appreciation to Dr. Rizwana Shaikh, our Head of Department who helped us in stimulating our efforts and knowledge received in these 4 years. Our admiration has no bounds towards our project guide Prof. Suvarna Chaurse who has guided to her fullest and has tried to inculcate all her experience and knowledge into us.

VII. REFERENCES

- [1]G. Preethi, P. Venkata Krishna, Mohammad S. Obaidat, V. Saritha, Sumanth Yenduri."Application of Deep Learning to Sentiment Analysis for recommender system on cloud,"in 2017 International Conference on Computer, Information and Telecommunication Systems (CITS),Dalian, China, July 2017.
- [2]Marco Pota, Massimo Esposito, Marco A. Palomino, Giovanni L. Masala." A Subword - Based Deep Learning Approach for Sentiment Analysis of Political Tweets," in 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA),Krakow, Poland, May 2018.
- [3]Dan Li,Jiang Qian. "Text sentiment analysis based on long short-term memory" ,in 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI),Wuhan, China,December 2016.
- [4]Saurabh Dorle, Dr. Nitin Pise. "Political Sentiment Analysis through Social Media" ,in 2018 Second International Conference on Computing Methodologies and Communication (ICCMC),Erode, India, October 2018.
- [5]Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen, Muhammad Kamran, Babar Hayat, A. Rehman." Sentiment Analysis Using Deep Learning Techniques: A Review", in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.
- [6]Amandeep Kaur, Deepesh Khaneja, Khushboo Vyas, Ranjit Singh Saini. Report on "Sentiment Analysis on Twitter using Apache Spark", SYSC 5807-Advanced Topics In Computer Systems, October 2017.
- [7]Imran Sheikh, Irina Ilina, Dominique Fohr. "Segmentation and Classification of Opinions with Recurrent Neural Networks". IEEE Information Systems and Economic Intelligence, May 2017, Al Hoceima,Morocco. 2017.
- [8]Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin. "Sentiment Analysis Techniques in Recent Works" ,Science and Information Conference 2015,London, UK, July 2015.
- [9]Nadheesh Jihan, Yasas Senarath, Surangika Ranathunga. "Aspect Extraction from Customer Reviews Using Convolutional Neural Networks", International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, September 2018.

[10]Alson Cahyadi, Masayu Leylia Khodra. “Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory”, 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), Krabi, Thailand August 2018.

