**A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis**

1. Introduction

Aspect-level opinion mining establishes a relation between different aspects of an item and its polarity. An aspect of a product means an attribute or feature of a product. For sentiment analysis, the identification of aspects is a very important issue. There are two types of aspect, explicit aspect, and implicit aspect.

1. Related Work
2. Sentiment Analysis Methodology Background:
   1. Different Levels of Sentiment Analysis

3.1.1. Document Level:

In Document Level sentiment analysis, it is analyzed whether the document expresses a positive or negative sentiment.

3.1.2. Sentence Level:

In Sentence Level sentiment analysis, the document is broken into some sentences and each sentence is treated as a single entity and analyzed at a time.

3.1.3. Aspect Level:

In Aspect Level, the main task is to extract aspect terms of the product and then customer feedbacks are analyzed on the basis of the extracted aspects.

* 1. Parts of Speech (POS) Tagging

Parts of Speech (POS) tagging is a form of annotating text and each word is a tag with Parts of Speech. Tokens are marked

with their corresponding word by the POS Tagger. Part-of- Speech tags are assigned to character strings. Each sentence can be categorized into a group of determiners, verbs, nouns, etc.

* 1. Dependency Parsing

The grammatical structure of a sentence and the relationships between ‘‘Main” words and the word which modify those main words can be obtained through a dependency parser.

* 1. Cluster Analysis

Cluster analysis is required in text mining for making a group of objects. After POS tagging and dependency parsing, lots of aspects are collected. To increase accuracy, aspects are categorized with the predefined set of aspects using hierarchical clustering.

* 1. Convolutional Neural Network (CNN) for Text Classification
     1. Word Embedding

Word embedding is a method where words or phrases from the vocabulary are mapped to vectors of real numbers. All the words in the input sentence are encoded as word vector.

* + 1. Convolution
    2. Pooling
    3. Dropout Regularization
    4. Fully-connection
    5. Softmax layer

1. Material and Methods
   1. Data Collection
   2. Data Pre-processing Method

The pre-processing method is used to clean data and convert data in the proper format for further analysis.

(i) Remove URL link: URL links do not carry much information regarding the sentiment of the tweet. So links are removed from tweets. (ii) Remove numbers: Generally, numbers have no use for measuring sentiment and are removed from the tweets in order to refine the tweet content. (iii) Convert acronyms: Acronyms are ill-formed words and are common in tweets. So acronyms are replaced by the original words through acronym dictionary. (iv) Words in tweets that contain repeated letters have been converted to their original English form. Words with repeated letters, e.g. ‘‘coooool” are replaced by ‘‘cool”. (v) Unnecessary white spaces and tabs are removed. (vi) All tweets are converted to lower case. All these steps are performed using R software with utils package.

4.3. Part-of-Speech(POS) Tagging

4.4. Modification of existing method by the implementation of some rule-based approach.

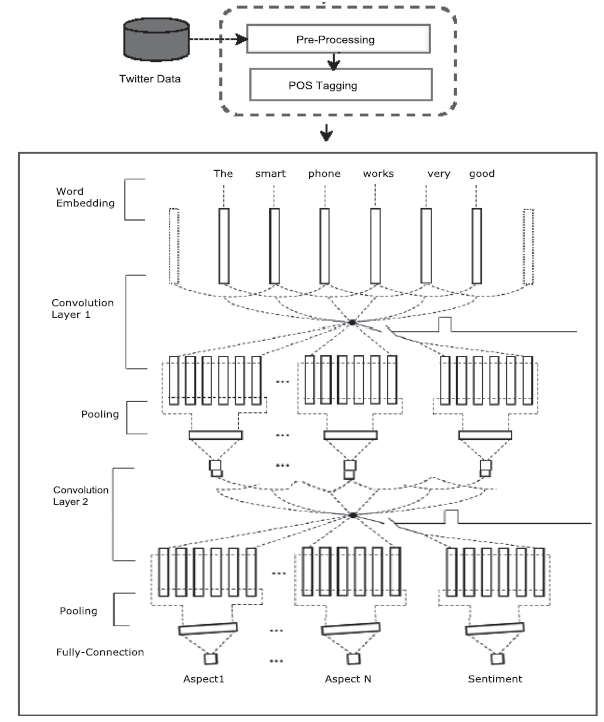
4.4.1. Dependency Parsing

4.4.2. Aspect Category Determination

4.5. Aspect Extraction with Convolutional Neural Network (CNN) and some rule-based approach

4.5.1. Word Embeddings

4.5.2. Proposed Network Architecture



4.6. Rule-Based Approaches

4.6.1. Rule-Based Aspect Extraction Approach

4.6.2. Rule-Based Sentiment Evaluation Approach

4.6.3. Handling Negation Word

4.7. Training set and Testing set Making

4.7.1. Initializing CNN

4.8. Package Used

5. Results & Analysis

6. Conclusion