Opinion Phrase Extraction and Aspect-Product Prediction from Reviews using Deep Learning

Maria Ann Toms
PSID - 2262086
Department of NSM
University of Houston
mtoms@cougarnet.uh.edu

Abstract

In the age of online shopping, customer reviews play a vital role in shaping purchasing decisions. By using advanced machine learning techniques, the proposed approach identifies opinion phrases and predicts which aspects and products customers are concerned about. Knowing the main opinion from the product reviews along with the aspect and the corresponding product details is crucial for companies to understand what customers like or dislike about their products. Leveraging Bidirectional Long Short-Term Memory (BiLSTM) for opinion phrase extraction and Linear SVC with TF-IDF features for aspect and product prediction, the methodology addresses the complexities of customer review analysis. Extensive experiments on a diverse dataset demonstrate the effectiveness of the proposed approach in extracting opinion phrases and predicting aspects and products with high precision and recall. The methodology contributes to advancing sentiment analysis and opinion mining techniques, offering businesses valuable insights for product improvement and decision-making.

1 Introduction

In the phenomenal growth of online shopping, people rely heavily on the reviews and experiences shared by others to decide what to buy. These customer reviews give us a lot of information about products, services, and how people feel about them. But with so many reviews out there, it's hard for businesses and customers to go through them all and understand what they mean. This is where automated methods come in handy. Opinion mining and sentiment analysis are techniques that help us understand the feelings expressed in these reviews.

By extracting opinion phrases and identifying the associated aspects and products, businesses can gain valuable insights into customer preferences, opinions, and satisfaction levels. This information can be used to improve products, enhance customer service, and make data-driven decisions. Moreover, customers can benefit from summarized opinions and sentiments, enabling them to quickly grasp the overall perception of a product or service. Traditional rule-based approaches and machine learning techniques such as Conditional Random Fields (CRF) have limitations in capturing the nuances and dependencies within the text. Deep learning models, particularly Recurrent Neural Networks (RNN) and their variants, have shown promise in sequence labeling tasks. However, there is still room for improvement in terms of accuracy and efficiency in opinion phrase extraction and aspect-product prediction.

The main objective of this paper is to propose a comprehensive machine learning approach for extracting opinion phrases and predicting aspects and products from customer reviews. I aim to address the limitations of existing methods by leveraging the power of deep learning and traditional machine learning techniques. My contributions can be summarized as follows:

1. I propose a Bidirectional Long Short-Term Memory (BiLSTM) model for opinion phrase extraction. The BiLSTM architecture captures the contextual information and dependencies within the review sentences, enabling accurate sequence labeling of opinion phrases.

2. I employ Linear SVC with TF-IDF features for aspect and product prediction. This approach allows me to effectively classify the extracted opinion phrases into their corresponding aspects and products.

3. I conduct extensive experiments on a dataset of customer reviews and evaluate the performance of my proposed approach using stan-

dard metrics such as precision, recall, F1-score, and accuracy. The results demonstrate the effectiveness of my model in extracting opinion phrases and predicting aspects and products.

4. I provide a detailed analysis of the results and discuss the insights gained from the experiments.

By addressing the challenges in opinion phrase extraction and aspect-product prediction, my work contributes to the advancement of sentiment analysis and opinion mining techniques. The proposed approach can assist businesses in extracting valuable insights from customer reviews, enabling them to make informed decisions and improve customer satisfaction.

2 Related work

The paper "Extracting Aspect-Specific Sentiment Expressions Implying Negative Opinions" [1] outlines the groundwork for this study. This research tackles the challenge of extracting precise phrases from customer reviews that express negative opinions about specific product aspects (e.g., "battery dies quickly"). This type of fine-grained analysis goes beyond identifying general positive or negative sentiment. By pinpointing exact issue areas, companies can make targeted product improvements, and customers can make informed decisions based on detailed feedback. The work focuses on two primary tasks: Task I - Issue Sentence Classification and Task II - Phrase Boundary Identification.

Task I aims to identify if a sentence with a known aspect word expresses a negative issue related to that aspect. To achieve this, researchers manually labeled sentences from Amazon reviews (1-3 stars) to create a custom dataset with issue phrases. They used various features for classification, including basic word and part-of-speech n-grams, pivot features based on the aspect, and latent semantic features from the ME-ASM model. The SVM classifier, with 5-fold cross-validation, proved most effective, especially when combining all feature types.

Task II addresses identifying the boundaries of issue phrases within sentences. This task used both unsupervised heuristic baselines, with predefined rules, and more complex sequence modeling approaches. Conditional Random Fields (CRFs) and their variants outperformed the heuristic baseline. The CRF-PSC model, which included phrase structural constraints, achieved the highest F1-score, highlighting the importance of recognizing valid patterns in issue phrases.

On experimenting with opinion mining and sentiment analysis, the focus was on classifying the overall sentiment of a text. However, later, the emphasis shifted towards identifying the sentiment expressed towards specific aspects or features of a product or service. One of the early approaches for aspect-based sentiment analysis involved extracting frequent nouns and noun phrases as potential aspects and determining the sentiment associated with each aspect using adjective-based opinion words. This approach laid the groundwork for many subsequent studies in the field.

Aspect and product prediction involve identifying the specific aspects or products mentioned in a given text. Early approaches relied on frequency-based methods, which extracted frequent nouns and noun phrases as potential aspects. Topic modeling techniques, such as Latent Dirichlet Allocation (LDA), have been applied to aspect discovery and extraction, aiming to uncover the hidden topics or aspects discussed in a collection of documents. Supervised learning approaches, such as Support Vector Machine (SVM), have also been used for aspect and product prediction, utilizing various linguistic features, including word embeddings and dependency relations, to improve the performance of the model.

3 Data

3.1 Dataset Description

The dataset comprises customer reviews spanning a variety of products, such as earphones, GPS devices, keyboards, mouse, MP3 players, and routers. The dataset is carefully structured in a hierarchy, with separate folders indicating various product categories. Within each product folder, there are sub-folders denoting specific aspects or components of the products, like cord, jack, or wire. Moreover, individual text files within these aspect sub-folders contain reviews that focus on particular aspects.

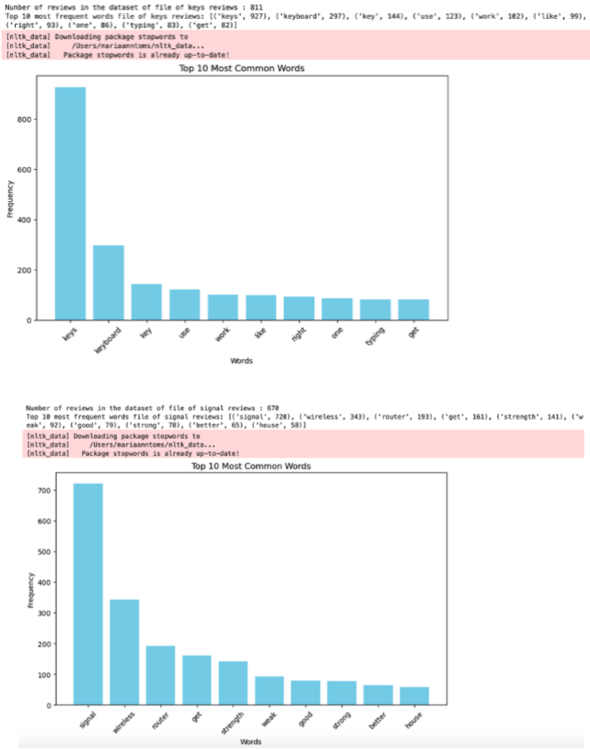

Figure 1: Histogram to visualize the top most frequent words

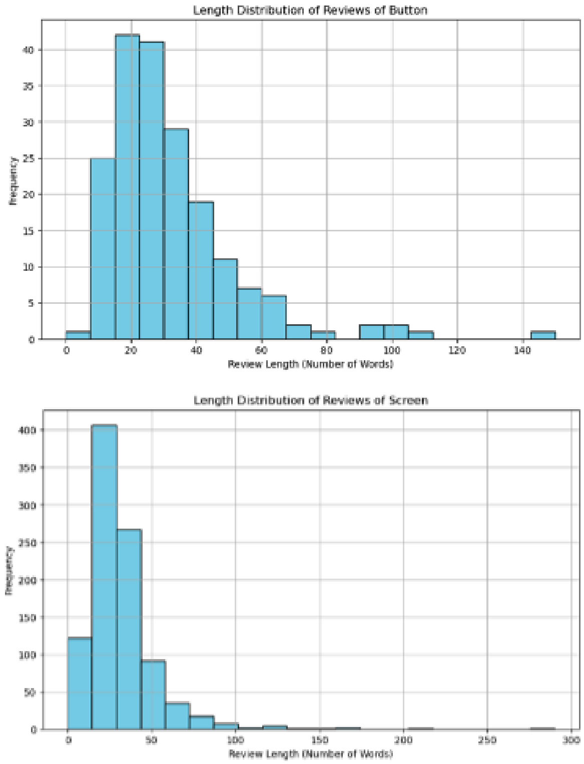

Figure 2: Histogram to visualize the spread of review lengths

3.2 Data Analysis

Counting the total number of reviews for a particular product, identifying the most frequent words in the dataset while excluding common stop words, and creating a histogram to visualize the distribution of the top most frequent words. This comprehensive approach allows for a deeper understanding of the text content, from quantifying the volume of reviews to visualizing the prominence of specific words within the dataset. The outputs for some of the products keyboard and router are given in figure 1.

The histogram provides insights into the spread of review lengths across the dataset, aiding in understanding the variability and typical length of reviews. The output for few of the files are given in figure 2.

Performing sentiment analysis on reviews stored in text files organized within a folder structure. It categorizes each review's sentiment as positive, negative, or neutral based on predefined sets of positive and negative keywords. The sentiment analysis is conducted for each aspect of the product, with the aspect names derived from the file names. The output for the different products are as given in the figure 3.

4 Methodology

4.1 Data Collection and Pre-processing

The dataset is stored in a CSV file format and loaded using the pandas library. It contains a column named 'Reviews' that holds the text of the product reviews. There is another column called 'OpinionPhrases' that contains the opinion or the main focus point in the review. The challenging part of the model would be to identify the opinion phrase when a new review is given to the model. The 'Reviews' column undergoes tokenization, which is the process of splitting each review into individual words or tokens. This step is crucial for further processing and analysis of the text data. A vocabulary of unique words is constructed from the tokenized reviews. This vocabulary represents the set of distinct words present in the entire dataset and is used to determine the size of the vocabulary for subsequent steps.

4.2 Feature Extraction and Representation

Word embeddings are used to convert the textual data into numerical representations that can be processed by machine learning models. They capture semantic and syntactic relationships between

```
Aspect-Sentiment Matrix for mp3_player :
Aspect       Positive    Negative    Neutral
interface    6           3           22
button       16          5           169
jack         6           3           33
screen       9           1           81
```

```
Aspect-Sentiment Matrix for earphone :
Aspect       Positive    Negative    Neutral
wire         34          8           182
cord         80          9           280
jack         9           3           74
```

```
Aspect-Sentiment Matrix for gps :
Aspect       Positive    Negative    Neutral
direction    62          27          380
voice        56          19          263
software     32          13          269
screen       159         46          753
```

```
Aspect-Sentiment Matrix for keyboard :
Aspect       Positive    Negative    Neutral
pad          80          13          250
keys         135         31          645
range        41          15          110
spacebar     16          11          103
```

```
Aspect-Sentiment Matrix for mouse :
Aspect       Positive    Negative    Neutral
wheel        161         21          629
pointer      9           2           42
battery      147         32          405
button       206         23          812
```

```
Aspect-Sentiment Matrix for router :
Aspect       Positive    Negative    Neutral
signal       90          38          542
wireless     170         55          1938
connection   114         40          1033
firmware     67          30          946
```

Figure 3: Consolidated sentiment analysis report of the reviews

```
0       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
1       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
2       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
3       [0, 0, 0, 0, 0, B-OPINION, I-OPINION, I-OPINIO...
4       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
                                ...
3536    [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, B-OPIN...
3537    [0, 0, 0, 0, 0, 0, 0, B-OPINION, I-OPINION, I-...
3538    [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
3539    [0, 0, B-OPINION, I-OPINION, I-OPINION, 0, 0, ...
3540    [0, B-OPINION, I-OPINION, I-OPINION, I-OPINION...
Name: Opinion_Labels, Length: 3541, dtype: object
```

Figure 4: Labels assigned for the tokens

words. The Keras Tokenizer is employed to fit the tokenized reviews and convert them into sequences of integers. Each unique word in the vocabulary is assigned a corresponding integer value. The sequences are padded to a fixed length of 100 using post padding. This step ensures that all sequences have the same length, which is necessary for feeding them into the neural network model. Post padding adds zeros at the end of shorter sequences to match the desired length.

4.3 Feature Extraction and Representation

Generating word-level labels for opinion phrases in each review. The labeling scheme follows the BIO (Beginning-Inside-Outside) format, where 'B-OPINION' denotes the beginning of an opinion phrase, 'I-OPINION' represents the inside of an opinion phrase, and 'O' indicates non-opinion words. The labels are converted into numerical representations using a label mapping dictionary. This mapping assigns unique integer values to each label class, enabling the model to process the labels as numerical data. The opinion labels generated for the tokens is as given in figure 4:

4.4 Model Architecture and Training

1. Neural Network Architecture

    - Input Layer: The input layer accepts the padded sequences of reviews. Each review is represented as a sequence of integers, where each integer corresponds to a word in the vocabulary.
    - Embedding Layer: The embedding layer converts the integer representations of words into dense vector representations. These embeddings capture the semantic relationships between words and allow the model to learn meaningful representations.

- Bidirectional LSTM Layer: The bidirectional LSTM (Long Short-Term Memory) layer is used to capture the sequential information and long-term dependencies in both forward and backward directions. It helps the model understand the context and capture relevant information from the entire review.
- Dropout Layer: The dropout layer applies regularization to prevent overfitting. It randomly sets a fraction of the input units to zero during training, which helps the model generalize better to unseen data.
- Output Layer: The output layer produces the predicted labels for each word in the review using a softmax activation function. The softmax function converts the model's outputs into probability distributions over the label classes.
- Model Compilation: The model is compiled with an Adam optimizer, which is an adaptive learning rate optimization algorithm. The loss function used is sparse categorical cross-entropy, which is suitable for multi-class classification tasks. The accuracy metric is used to evaluate the model's performance during training.

2. K-Fold Cross-Validation K-fold cross-validation is employed to assess the model's performance and generalization ability. It helps to estimate how well the model will perform on unseen data. The data is split into k=5 folds. For each fold, the data is divided into train, validation, and test sets. The model is trained on the train set, validated on the validation set to monitor its performance and tune hyper parameters, and finally evaluated on the test set to assess its performance on unseen data. For each fold, evaluation metrics such as accuracy, precision, recall, and F1-score are calculated. These metrics provide insights into the model's performance in correctly identifying opinion phrases.

4.5 Model Evaluation and Visualization

- Test Set Evaluation: The trained model is evaluated on the test set in each fold. This step assesses how well the model generalizes to unseen data and provides an unbiased estimate of its performance.
- Prediction and Comparison: Predictions are made on the test set using the trained model. The predicted labels are compared with the true labels to calculate evaluation metrics.
- Confusion Matrix: A confusion matrix is plotted to visualize the model's performance in classifying the opinion phrases. It provides a tabular summary of the model's predictions, showing the counts of true positive, true negative, false positive, and false negative predictions for each class.
- Training and Validation Metrics: The training and validation loss and accuracy are plotted over the epochs. These plots help to monitor the model's learning progress, identify any overfitting or underfitting, and determine the optimal number of training epochs.

4.6 Aspect and Product Prediction

TF-IDF (Term Frequency-Inverse Document Frequency) vectorizers are used to extract features from the reviews for aspect and product prediction. TF-IDF assigns weights to words based on their frequency in individual reviews and their rarity across the entire dataset, capturing the importance of each word. Linear SVC classifiers are trained on the extracted TF-IDF features for aspect and product prediction.

This methodology highlights the comprehensive approach employed for opinion mining and aspect-based sentiment analysis. It covers data preprocessing, feature extraction using word embeddings and TF-IDF, model architecture design with a bidirectional LSTM neural network, training using k-fold cross-validation, evaluation using various metrics and visualizations, and inference for opinion phrase, aspect, and product prediction.

The combination of deep learning techniques for opinion phrase extraction and traditional machine learning methods for aspect and product prediction showcases a hybrid approach to tackle the complex task of understanding and analyzing sentiments expressed in product reviews.

5 Results

The confusion matrix and both the plots for training, validation, testing loss and the plots for

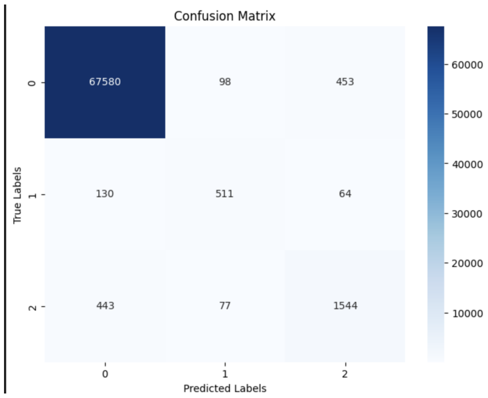

Figure 5: Confusion Matrix of fold 1

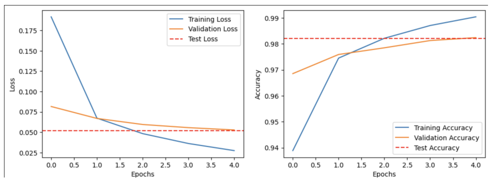

Figure 6: Plotting Loss and Accuracy of fold 1

training, validation, testing accuracy for all the 5 folds are as given in figures 5-14.

During the training process, the model's loss and accuracy are monitored on both the training set and the validation set. The training loss and accuracy indicate how well the model is learning from the training data, while the validation loss and accuracy provide insights into the model's performance on unseen data. Plotting the training and validation loss/accuracy over the epochs helps to assess the model's learning progress and identify any overfitting or underfitting.

The average accuracy of 0.982 indicates that the model correctly predicts the opinion phrases, aspects, and products. The average precision of 0.982. The average recall of 0.982. The average F1-score of 0.982 is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. The consistency of the average accuracy, precision, recall, and F1-score values at 0.982 suggests that the model is performing exceptionally well across all the evaluation metrics. This level of performance indicates that the model is highly effective in opinion mining and aspect-based analysis tasks.

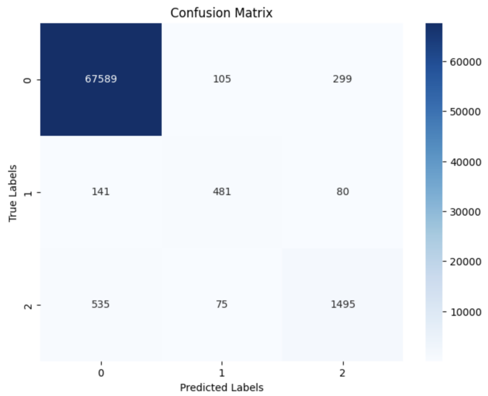

Figure 7: Confusion Matrix of fold 2

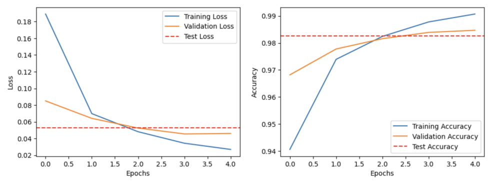

Figure 8: Plotting Loss and Accuracy of fold 2

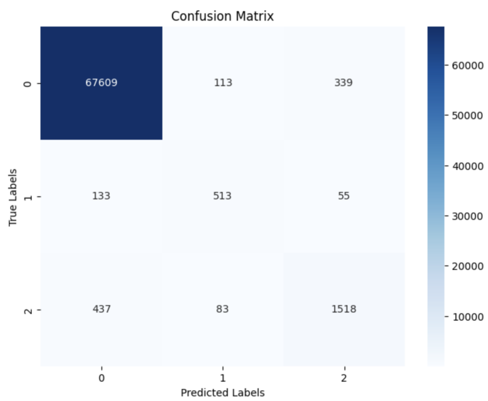

Figure 9: Confusion Matrix of fold 3

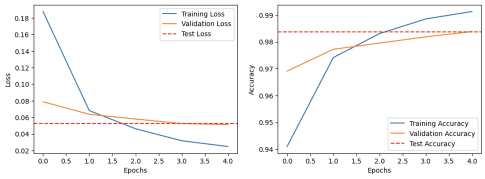

Figure 10: Plotting Loss and Accuracy of fold 3

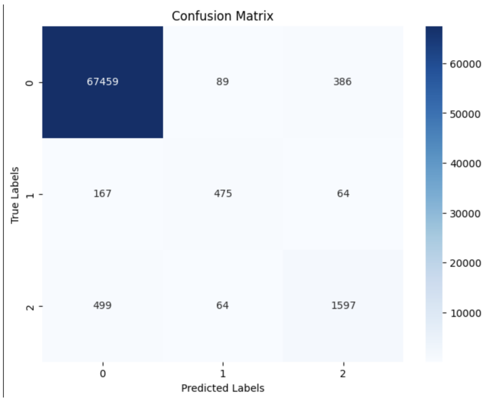

Figure 11: Confusion Matrix of fold 4

6 Conclusion

In this project, we developed and applied a deep learning method for opinion mining and aspect-based sentiment analysis. Using customer review data, we converted text into numerical representations with word embeddings. A bidirectional LSTM model was trained to extract opinion phrases and identify associated aspects and products within reviews. The model was thoroughly tested with k-fold cross-validation to ensure reliable results.

Although the results were encouraging, there are several areas for further development and enhancement to improve the model's performance and broaden its use:

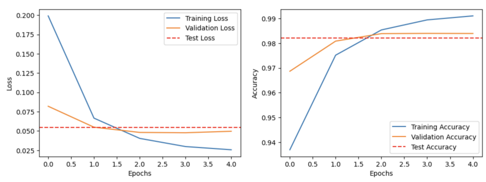

Figure 12: Plotting Loss and Accuracy of fold 4

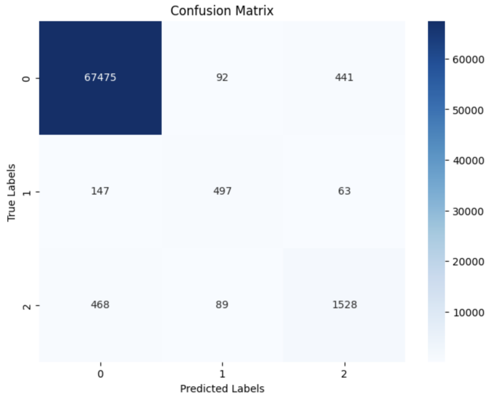

Figure 13: Confusion Matrix of fold 5

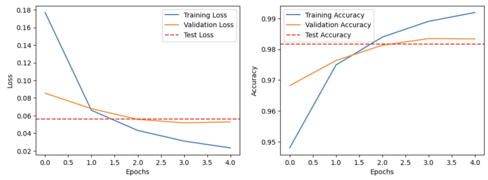

Figure 14: Plotting Loss and Accuracy of fold 5

1. Advanced Language Models: Implementing modern contextual word embeddings like BERT or ELMo could improve semantic understanding and enhance the model's accuracy in capturing sentiment from reviews.

2. Complex Linguistic Phenomena: Developing advanced methods to manage negation, intensifiers, and other complex language features could lead to more precise sentiment analysis.

3. Adapting to Different Domains and Languages: Examining how well the model performs across various domains and languages could extend its applicability, making it useful in a wider array of sentiment analysis contexts.

4. Sentiment-Based Recommendation Systems: Integrating the sentiment analysis with Recommendation systems could transform personalized recommendations by taking into account users' sentiments and preferences, potentially increasing user satisfaction and engagement.

References

[1] Mukherjee, A., 2018. Extracting aspect specific sentiment expressions implying negative opinions. In Computational Linguistics and Intelligent

Text Processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, Revised Selected Papers, Part II 17 (pp. 194-210). Springer International Publishing.

[2] Chen, Y., 2015. Convolutional neural network for sentence classification (Master's thesis, University of Waterloo).

[3] Tang, D., Qin, B. and Liu, T., 2015, September. Document modeling with gated recurrent neural network for sentiment classification. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432).

[4] Lafferty, J., McCallum, A. and Pereira, F., 2001, June. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Icml (Vol. 1, No. 2, p. 3).

[5] Poria, S., Cambria, E. and Gelbukh, A., 2016. Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, pp.42-49.