Deep Learning to Improve Sentiment Analysis

Author

Rabia Noureen

FALL 2015-MS-15(CSE) 00000119941

MS-15 (CSE)

Supervisor

Dr. Usman Qamar

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

SEPTEMBER, 2019

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. The thesis is also according to the format given by the university.

Signature of Student

Rabia Noureen

Registration Number

00000119941

Signature of Supervisor

Usman Qamar

COPYRIGHT STATEMENT

ACKNOWLEDGEMENTS

I am thankful to Allah Almighty for giving me countless blessings and guidance in this research work. It was all possible because of His help. I want to thank my teachers **Dr. Usman Qamar**, **Dr. Farhan Khan** and **Ma'am Saba Bashir** for supporting me throughout the whole work and giving me ideas to improve my research. I couldn't have done this without them. They are the source of great knowledge and their expertise helped a lot in achieving my goals. I also want to mention the efforts of **Dr. Muhammad Abbas** and **Dr. Wasi Haider Butt** for guidance and support. Then I want to thank and appreciate efforts of my parents, siblings, and friends who encouraged me throughout the whole degree and made the things easy for me with their motivational assistance. I would like to express my gratitude to all the mentioned people and my department for assisting me in my publications.

Dedicated to my exceptional parents whose tremendous support and cooperation led me to this wonderful accomplishment.

ABSTRACT

The amount of unstructured data on the internet is increasing with each passing day. People express their opinions on social media, blogs, and forums in different ways. Therefore, the need to process that unstructured data grows as well. The processing is important for extracting knowledge patterns from the data available online. Sentiment analysis is a popular technique that is used for knowledge extraction from people's opinions. There are many businesses and product developers that use people's opinions as a basis for improvement. We have seen great progress in Natural language processing and Sentiment Analysis techniques over the past few years. However, traditional sentiment analysis approaches focused on some particular type of data by using machines learning models and failed to achieve the best performance in terms of accuracy. There are many short comings in previous studies. Therefore, a major challenge is to overcome these problems to extract useful data from the huge amount of data that is available online. This thesis has been conducted to reduce the research gap by coming up with a better solution to improve the accuracy of the existing models used for sentence level sentiment analysis tasks. Thus, we have proposed a neural network-based sequence model (RNN-LSTM) that is used for the sentiment classification from opinionated sentences. Our Sentiment classification model is based on two state of the art deep learning algorithms Recurrent Neural Network (RNN) and (LSTM). We have evaluated our approach on four sentiment classification datasets. Furthermore, we have also made a detailed comparison with popular baseline approaches. The results prove that the proposed technique acheives improved accuracy as compared to the existing models.

Key Words: Natural language processing, Sentiment analysis, Deep neural network, RNN, LSTM

TABLE OF CONTENTS

LIST OF FIGURES

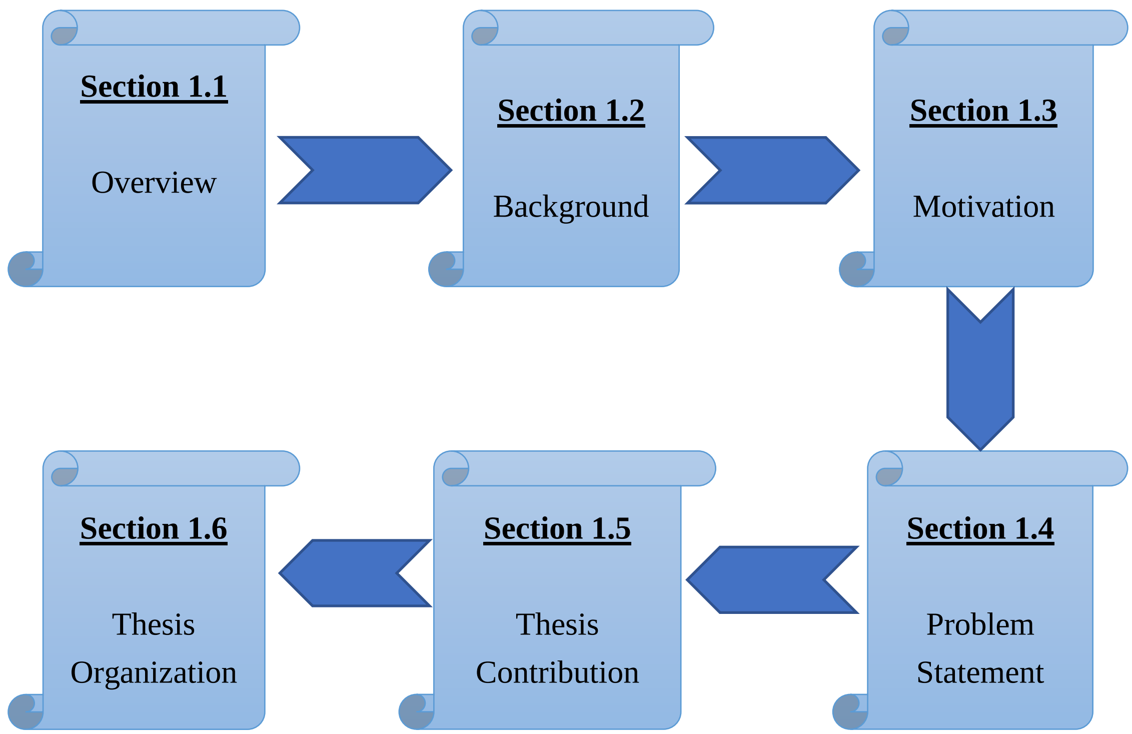
LIST OF TABLES

Chapter 1

Introduction

CHAPTER 1: INTRODUCTION

In this chapter detailed introduction of research is presented. **Section 1.1** shows the overview. **Section 1.2** discusses the background. **Section 1.3** contains the motivation for research work. **Section 1.4** describes the problem statement and **Section 1.5** and **Section 1.6** contains thesis contribution and thesis organization.



1.1 Overview

The amount of unstructured data over the internet is increasing day by day. As the data grows, likewise the need to intelligently process this unstructured data and extract accurate knowledge from it grows as well. We can see that there has been immense progress in the field of Natural Language Processing (NLP). Researchers have been working very hard to deliver efficient techniques and models that are able to perform better. These models are used to extract useful information from tons of unstructured data that is available online. These techniques have been used for sentiment analysis and document classification.

Moreover, NLP has a major goal of developing scalable and general models and algorithms. These algorithms are used to solve sentiment classification problems and learn useful features from the linguistic patterns. However, most of these approaches have their own short comings.

1.2 Background

Sentiment analysis is a popular field of research that deals with analyzing the sentiments, attitudes, opinions, emotions, and appraisals, targeted towards a specific attribute or entity that is specifically written in text. [1]. Social media is growing at a rapid pace and as a result of its growth people have been sharing their opinions in the form of forum discussions, news, reviews, comments, and blogs. The social media addiction has a large impact on society and business as a whole.

One of the main types that are focused these days is sentence-level sentiment analysis. Most of the research studies that have been conducted in this field were focused towards the polarity of opinionated sentences expressed online. This polarity is divided into three classes positive, negative and neutral. The textual content of sentences is analyzed to extract language clues. [2] [3]

[4]. These studies considered it as a general problem and didn't focus on the types of sentences. It is important to understand that the sentiments are expressed in different types of sentences. There are some sentences whose sentiment polarity is difficult to identify. It depends on the opinion targets that have been used in a sentence.

Furthermore, we cannot deny the fact that the opinionated sentences that are used for sentiment analysis can be expressed in a random and understated format. If we are just looking at each individual word, it would be very difficult to identify such sentences. We cannot generalize the sentiment analysis tasks by choosing a single technique that can be used for all sentiment analysis tasks [5]. Therefore, we need to deal separately with some specific types of sentences. Some models perform best in some situations while we may need to perform different operations on other types [1].

We have tried to address the short comings of existing models discussed in this thesis. Although these models perform best, still they are dealing with some specific types of datasets. Additionally, these models perform the best results based on the analysis done on some manually created features.

The inspiration for our new model came from analyzing the short comings of the existing models. The ideas from both the domains of deep learning and natural language processing. Deep learning is basically a field that has been derived from machine learning itself. This technique overcomes one of the major challenges in sentiment analysis. Deep learning basically automatically extracts and learns features representations from input data available in the raw form. The model then uses these features for addressing prediction and sentiment classification problems.

Many researchers have addressed the image classification problems by using deep learning techniques [6]. Some of them have used deep learning for solving speech recognition problems [7]. However, it is important to understand that the hierarchical and recursive structure of deep learning can help us to address problems in the sentence-level sentiment analysis.

This research study presents a new deep learning model that addresses the short comings in the previous researches. Based on the review of the existing work, we have designed a deep neural network model that is based on two state of the art models. We have combined the long short-term memory (LSTM) model with Recurrent neural network (RNN) model. The model that we have proposed is RNN-LSTM. We will be using opinionated sentences for extracting target expressions with the help of this model.

We will be using this deep learning model for sentiment classification based on four popular datasets. These datasets are CR, MR, SST-1, and SST-2. We will evaluate the effectiveness of this technique through comparative analysis based on existing studies.

1.3 Motivation

Today social media has established it self as an essential element of our life. The usage ratio of top social networking sites does not significantly change on a yearly basis. However, the level of usage varies in different countries. It is interesting to observe that the use of social media is diverging in some countries. According to the Global Digital Report 2019, the number of internet users around the world is 4.388 billion. The number is increasing by 9.1 % on a yearly basis. The second finding suggests that there are more than 3.484 billion social media users. The annual growth increases at the same rate.

Sentiment Analysis helps the businesses in perceiving their services or brand in the mind of the customers. You can take advantage of the technique to develop effective branding techniques and strategies that can boost your leads. They can measure the ROI of their marketing campaigns. Furthermore, sentiment analysis can help them to provide better customer support. However, it only depends on the tools that you use to take advantage of the valuable data that you have in the form of opinions. The existing sentiment analysis techniques failed to achieve best performance in terms of accuracy.

1.4 Thesis Objective and Contribution

We have conducted this research in order to find the best performing deep learning technique that can help in sentence-level sentiment analysis tasks. We have proposed a technique that outperforms all other techniques used in the past with higher accuracy. This research study has been performed keeping in mind the following objectives.

- To develop a novel deep learning framework that can be helpful in sentence level sentiment classification tasks.
- To identify sentiment analysis on datasets that are extracted from thousands of reviews available on different websites in an efficient and accurate manner. To validate our technique by comparing it with state-of-the-art models and approaches that have been used in the past.

Our proposed technique satisfies the above-mentioned objectives and uses an efficient technique that can be used to classify sentiments based on different types of data.

1.5 Thesis Outline

This research study has been conducted in order to overcome the short comings inn the existing researches who addressed the sentiment classification problem. A brief introduction of the problem has been presented in **CHAPTER 1.** This section includes the overview, background study, thesis motivation, objective and contribution of this thesis. The detailed literature review has been done in **CHAPTER 2.** This section discusses some of the major work that has been done in the past in the domain of sentiment analysis by using machines learning and deep learning models. The proposed methodology has been discussed in **CHAPTER 3.** Moreover, the implementation details of the proposed framework have been discussed in **CHAPTER 4.** We have briefly discussed the proposed technique along with results and discussions in **CHAPTER 5.** Finally, the research work has been concluded in **CHAPTER 6** and this section also discusses the future work**.**

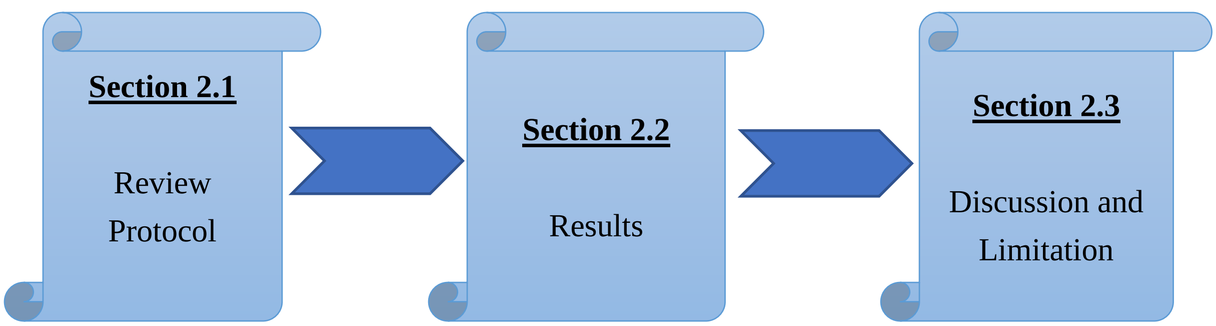
Chapter 2

Literature Review

CHAPTER 2: LITERATURE REVIEW

Social network analysis is the process of investigating social links through the use of networks. Aim of this SLR is to determine what are different categories of networks that have been recognized through social networks and also to identify what techniques so far has been proposed for inferring social relationships.

This chapter presents the literature review carried out for the research. **Section 2.1** discusses the review protocol. **Section 2.2** presents results obtained followed by **Section 2.3** presents the discussion and Limitation. **Section 2.4** describes the answer to RQ’s and Finally **Section 2.5** describes research gap.



2.1 Review Protocol Development

2.1.1. Inclusion and Exclusion Criteria

We have followed the following Inclusion and Exclusion criteria in order to conduct this research study:

1. We have just selected the research studies that are dealing with the text-based sentiment analysis. Most of these papers have used deep learning techniques to classify textual data. While others discuss a combinational of other models and techniques that have been used in the past. The selected research studies also answer the research questions that we are going to answer in this study. Moreover, we have rejected those research studies that didn't discuss textual sentiment analysis.

2. We have only selected those research studies that have been conducted in the timeline from 2010 to 2019.

3. We have downloaded the research studies from five famous digital libraries. These libraries are TAYLOR & FRANCIS, IEEE, ELSEVIER and SPRINGER and ACM. All of these digital libraries are known for publishing state of the art research studies. Details are provided in **Table 1**.

4. We have rejected all of the research studies that were redundant and focused on only those that contributed state of the art techniques in the existing research work.

2.1.2. Search Process

We have used the manual technique to complete the search process. The manual search ensured the primary search references. The preliminary stage included the downloading of research studies from Elsevier, IEEE, Taylor and Francis, ACM, and Springer. These digital libraries are known for providing the most notable conference proceedings and impact journals that cover the area of deep learning and sentiment analysis.

In order to made sure that we get relevant research papers for this review, we have used general keywords for the goal of finding the maximum number of papers as mentioned in **Table 2**.

We have used the AND operator with the keywords. The steps that we performed during the research process are shown in **Figure 1**.

Table 1. Details of research studies against each database

Sr.#	Scientific Database	Selected Research Works	No. of Researches
1	IEEE	[25] [26]	2
2	SPRINGER	[9] [18] [19] [20] [20] [21] [23]	7
3	ELSEVIER	[8] [10] [11] [12] [13] [14] [15] [24]	8
4	ACM	[16] [17]	2
5	Taylor and Francis	[22]	1

Table 2. Details Of Search Terms And Search Results

Sr.#	Search Terms	Operator	IEEE	Springer	ACM	ELSEVIER	Taylor and Francis
1	Deep Learning	AND	24499	2215	18756	29089	1040
2	Sentiment Analysis	AND	8965	3056	256	4859	2894
3	Deep neural network	AND	3801	8601	630	42534	1032
4	RNN	AND	10772	18851	8286	35273	962

2.1.3. Quality Assessment

The quality assurance and assessment are done in order to ensure that the selected papers are enough to formulate a decision. Therefore, in order to identify the value of deep learning models, the authenticity of the results in the specific researches. We have evaluated all the research studies and thus used the quality criteria to evaluate these publications.

Q1. Does analysis is linked with the issues discussed in the research studies.

Q2. Does the study explain the context of the analysis?

Q3. Did the research study define the review methodology?

Q4. Does the research study describe the data gathering procedure?

Q5. Does the study explain the data analysis procedure?

Hence, we have evaluated a total of 20 research studies keeping in mind the above-mentioned quality assurance criteria. These criteria define the credibility of the chosen studies.

The number of selected researches selected from 2010-2019 are shown in **Figure 1**.

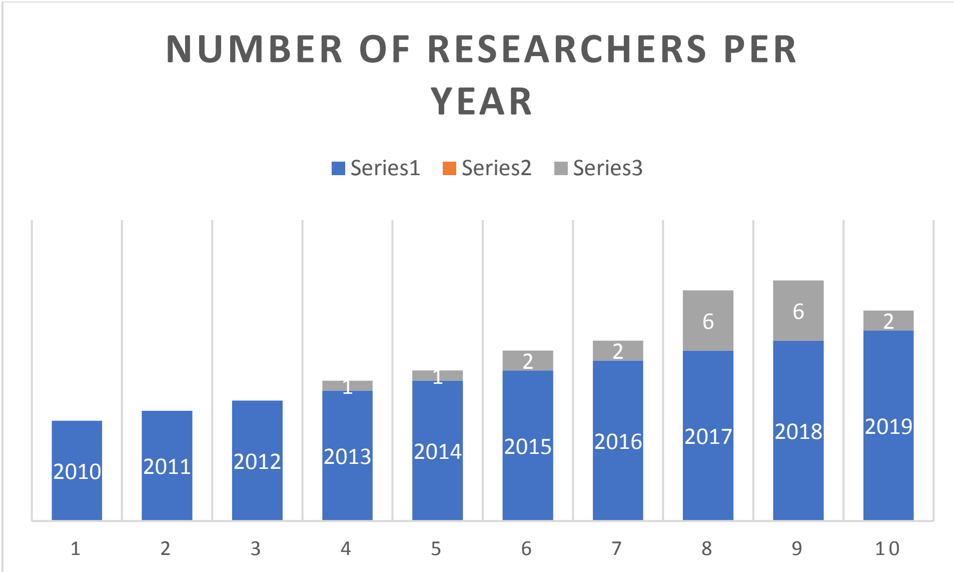


Figure 1. Number of Selected Researches (Yearly)

2.1.4. Data Extraction and Synthesis

We have done a detailed analysis of the research studies selected for the review process. Furthermore, it involves the data extraction and synthesis process in order to perform scrutiny of the issues covered in the research studies. The methods, approaches, models used for sentiment analysis.

Once we successfully extracted the information from the research studies, we started working on the analysis phase. The analysis phase was completed using quantitative and qualitative analysis techniques.

The data extraction and synthesis that we performed for our selected publications are shown in **Table 3**.

Table 3. Data Extraction and Synthesis

Sr. #	Description	Details
1	Bibliographic information	Author, Title, Publication Year, publisher details, and type of research (i.e. journal or conference)
Extraction of Data		
2	Overview	The major aim of conducting the study and what the research is about
3	Results	The results extracted from selected publications

4	Data Collection	Qualitative and quantitative methods we used for the collection of data
5	Validation	The techniques used to validate the outcomes of this research study
Synthesis of Data		
6	Identification of text-based machines learning models and approaches	The algorithms and models that are used for sentiment classification on textual data
7	Identification of machines learning models	The machines learning models that are used for sentiment classification
8	Identification of Deep Learning models	The deep learning models that are used for sentiment classification

2.2 Results

Table 1 shows an overview of the research studies we have selected from all the five scientific databases. We have obtained the answers of he Research Questions by conducting an exhaustive analysis of the chosen research studies. A brief overview of the results is provided in the section below.

This section discusses the results of the research study that are obtained to answer the questions asked in the systematics review. The research questions were formulated to analyze the existing techniques used for sentiment analysis. We focused on the following angles in this regard: sentiment analysis, model and framework, the accuracy that has been obtained to verify the methodologies.

We have included 20 publications based on our criteria after screening thousands of records. We have selected research studies that discussed different approaches used for textual sentiment analysis from 2010 to 2019

2.2.1 Deep Learning for Sentiment Analysis

We have identified a total of 11 deep learning models that have been used for sentiment analysis in the previous studies. **Table 4** shows the list of deep learning models that have been used for sentiment analysis.

Table 4. Identified Deep Learning Models

Algorithms		
Sr.#	**Name of Algorithms**	**References**
1	Deep belief nets (DNN)	[8]
2	RNN	[9] [24]
3	BiLSTM-CRF	[10]
4	CNN	[11] [20]
5	Deep belief networks (DBN)	[12]
6	CNN-LSTM	[14]
7	CNN+ Rule based	[15]
8	LSTM-CNN	[16]
9	CNN-GRU	[23]
10	RNTN	[26]
11	DCNN	[27]

2.3 Discussion and Limitation

Deep learning has shown good results in many different domains including natural language processing (NLP). Deep Learning is the best technique that can be used to address various problems in data analysis and learning problems. We have observed a shallow learning process with data mining approaches. The deep learning models use various layers to transform the input. The feature extraction and data representation are done with the help of hidden layers used in the Deep Learning process. The hierarchy-based learning process helps us efficiently perform sentiment analysis tasks with higher accuracy. All of the above-mentioned attributes suggest that Deep Learning is the perfect technique that can be used to perform sentiment analysis. The research founded that various deep learning methods, approaches, and algorithms are already being used to perform sentiment analysis. We have done an extensive comparative analysis of the models, methods, and approaches that have been used in different publications. The results are mentioned in **Table 5.** Most of the research studies have used publicly available data sets.

We have defined some parameters for the analysis and comparison of the selected techniques that have been used in the previous researches. The following parameters have been used for these purposes.

Research Problem

The research problem is the problem in the research domain that has been tackled in the study. Either the search has been done for sales analysis, economic review, predicting the product selection trend or to answer some research questions or hypothesis.

Proposed Approach

The proposed approach is basically the methodology or deep learning models that have been used to answer the research problem. It can either be any framework, process or model that has been used for sentiment analysis

Dataset

This section includes the datasets that have been used in order to conduct the research study. The data sets also propose the authenticity of the models. Some examples of datasets are Product Reviews, Movie Reviews, IMDB, Yelp, etc.

Validation Results

This section includes the validation results that have been used to validate the proposed approaches such as accuracy, F1-score, Absolute Error, and others.

Limitations

We have tried to ensure the accuracy and consistency of the research studies by adopting an exact search process. However, there is still a possibility that some major research studies have been missed out during this procedure.

Table 5. Comparative Analysis of Research Studies

Sr.#	Models	Datasets	Results	References
1	Deep belief nets (DBN)	Twitter database, Sina microblog (in Chinese), COAE2014 (in Chinese)	76%	[8]
2	RNN	Movie review, MPQA Opinion, Customer Review	86.7%,	[9]

3	BiLSTM-CRF	Movie Review, Customer Review, SST-1, SST-2	88.3%	[10]
4	CNN	Google Embeddings, Amazon Embeddings, SemEval 2014 dataset	87.17%	[11]
5	Deep belief networks (DBN)	Movie review (MOV), books (BOO), DVDs (DVD), electronics (ELE), and kitchen appliances (KIT).	74.9%	[12]
6	Ensemble model	SemEval2013/14, Vader, STS-Gold, Sentiment140, IMDB, PL04	86.49%	[13]
7	CNN-LSTM	Arabic Tweets	65.05%	[14]
8	CNN+ Rule based	Twitter, SST-1, SemEval Task 4	88.6%	[15]
9	LSTM-CNN	FiQA 2018	69%	[16]

10	Naïve Bayes, SVM	Textblob, SentiWordNet and Word Sense Disambiguation (WSD)	79%	[17]
11	CNN	Sentiment 140 corpus	74.9	[18]
12	NBM, SVM, MEM, DT	Turkish and English Movie and Product review	84.5	[19]
13	CNN-GRU	YNM, VLSP	85.8	[21]
14	RNN	Movie Review, MPQA, Customer Review	86.5	[22]
15	SVM-CNN-MKL	Multimodal Sentiment Analysis, Multimodal Emotion Analysis datasets	88.6	[23]
16	RNTN	Sentiment Treebank	80.7	[24]

17	DCNN	Tweet dataset	80.69	[25]

2.4 Answer to RQ'S

RQ 1: What are the algorithms and models that have been used for textual sentiment classification since 2010?

Answer: We have selected total of 20 publications based on the criteria that have already been mentioned in Section 2. We have done extensive analysis and study and evaluated that all of them presented algorithms and models have been used for textual sentiment analysis since 2010.

RQ 2: How many publications used machine learning models to classify sentiments from text since 2010?

Answer: The screening and analysis of the 20 research studies indicated that we have identified 9 machine learning models that have been used for sentiment analysis.

RQ 3: How many approaches and methods have been proposed to classify sentiments from deep learning models since 2010?

Answer: We have selected total of 20 research studies and identified that a total 11 publications used deep learning models for sentiment classification.

2.5 Research Gap

The volume of the data available on the internet is increasing day by day. There are many platforms that are being used by billions of people to express their opinions. Therefore, it is very difficult to analyses sentiments from the data available on the internet. There are many different techniques and models that have been used in order to predict sentiments from opinions. In this research study, we have analyzed the techniques that have been used for the sentiment analysis. All of these techniques have their own pros and cons and were successful in proving their proposed hypothesis. However, we cannot deny the fact that a major research gap still exists in the data collection methods that have been used in these approaches. Moreover, the existing deep learning models need to be improved to improve the accuracy of results. Most of the researchers have used individual deep learning models such as CNN that failed to outperform on some data sets. Therefore, a model is needed that can produces a higher accuracy score on sentiment datasets.

Figure 2 shows the distribution of various models used in the previous researches.

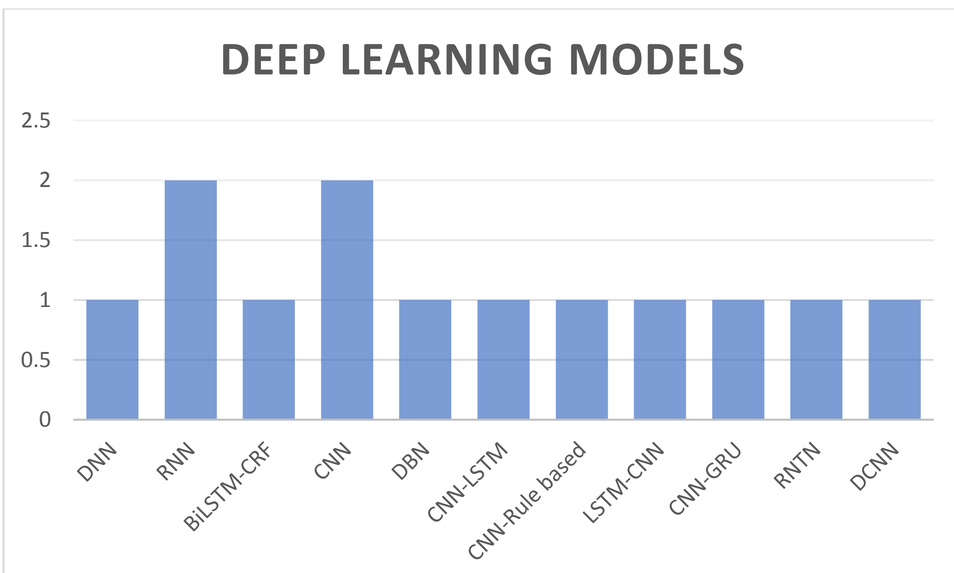


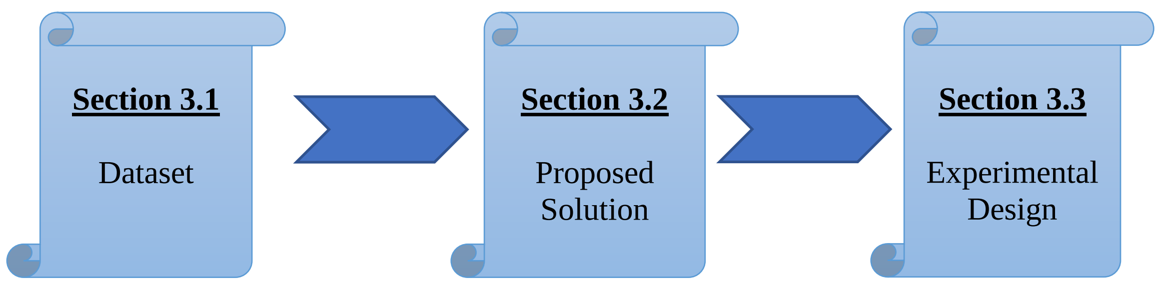
Figure 2: Deep Learning Models Used in The Previous Studies

Chapter 3

Proposed Methodology

CHAPTER 3: PROPOSED METHODOLOGY

The proposed methodology that has been used in this research is discussed in this section. **Section 3.1** lists the datasets that have been used in this research for extensive experimentation. We have discussed the details of the proposed framework in **Section 3.2**. Finally, we have presented the experimental design in **Section 3.3**.



3.1 Proposed Methodology

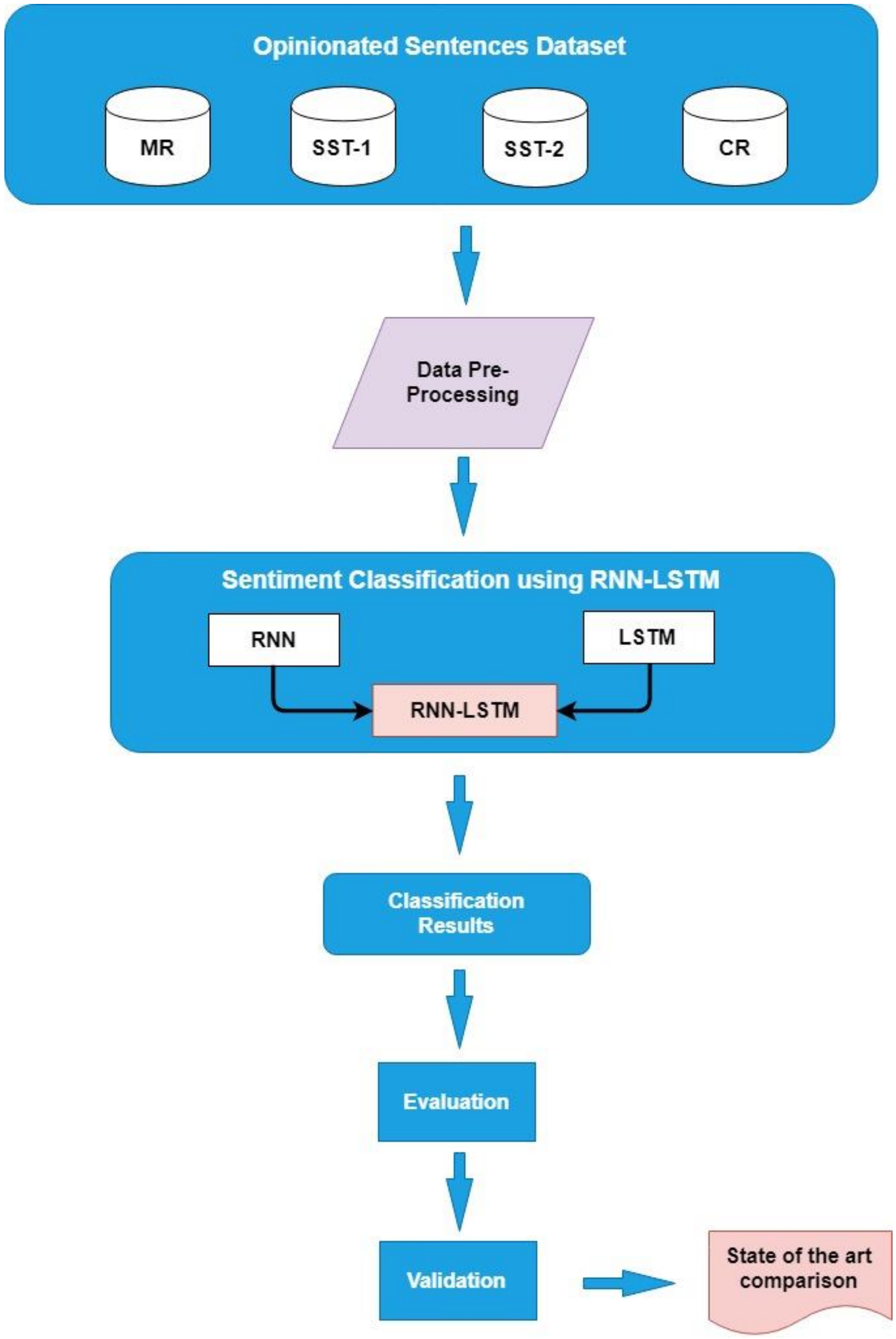


Figure 3. The framework of Deep Learning Model (RNN-LSTM)

3.2 Dataset

In order to train the RNN-LSTM model for we have used four publicly available datasets. We have also tested our framework on the following datasets:

1. **MR:** The first datasets that we have used is the Movie review sentence polarity dataset. There are different versions available at the moment and we have used v1.0. The number of positive and negative snippets included in the dataset are 5331 and 5331 respectively. The researchers have used Rotten Tomatoes to extract negative snippets. Moreover, positive snippets are extracted from 345 web site pages. Only those review were selected that have were marked as fresh. We have used 10-fold cross validation for testing.
2. **SST-1:** The second dataset that we have used in our research is Stanford sentiment treebank. This dataset has a total of 11855 sentences. All of these sentences are extracted from files obtained from the original pool of Rotten Tomatoes page. The sentences have five types of labels (very negative, negative, neutral, positive, very positive).
3. **SST-2:** The third dataset is extracted from the Stanford sentiment treebank with binary labels. We have removed neutral reviews from the SST-1 dataset. Moreover, the positive label is allocated to a very positive and positive datasets. Likewise, a negative label is allocated to the very negative and negative reviews. This dataset has 9613 sentences in total.
4. **CR:** Customer Review dataset is the final one that we have used for experimentation. It contains the reviews of 5 digital products and it is basically obtained from Amazon.com. The CR dataset contains 3771 sentences in total and these reviews are divided into 2405 positive reviews and 1366 negative reviews.

Table 7: Details of Datasets

Dataset	References	Research Studies
Movie Review	https://www.cs.cornell.edu/people/pabo/movie-review-data/	Pang and Lee (2005)
SST-1	http://nlp.stanford.edu/sentiment/	Socher et al. (2013)
SST-2	Same as SST-1 but with neutral reviews removed and binary labels	Kim (2014)
Customer Review	http://liu.cs.uic.edu/download/data/	Hu and Liu (2004)

Table 8: Dataset Attributes

Dataset	Classes	Labels	Sentences	Average Sentence Length
MR	2	Positive/Negative	10662	20
SST-1	5	Very Negative, Negative, Neutral, Positive, Very Positive	11855	18
SST-2	2	Positive/Negative	9613	19
CR	2	Positive/Negative	3771	19

3.3 Proposed Solution

3.3.1 Dataset Selection

Dataset selection is the first process in this framework. The proposed methodology requires a pool of opinionated sentences. As mentioned before, we have obtained four publicly available datasets MR, SST-1, SST-2, and CR.

3.3.2 Data Preprocessing

Now at the second step, the preprocessing of data is done. This step involves some processes that are required before we move ahead to the classification phase. It is basically a process where the dataset is prepared and cleaned at the initial level. There are many preprocessing tasks involved in the preprocessing phase. This step is important for noise reduction, performance improvement of the classifier. Furthermore, it also speeds up the entire classification process.

3.3.3 Sentiment Classification

The next step is to apply the sentiment classification algorithm (RNN-LSTM) on the opinionated datasets. Both of them have shown the best results in the previous results in the past. Our model RNN-LSTM performs classification of the sentiments. The results and then saved for the analysis phase.

3.3.4 Results Validation

Then the same process is repeated for all the datasets and compared against each other. The comparison shows which of the datasets perform better on the proposed approach.

3.3.5 State of Art Comparison

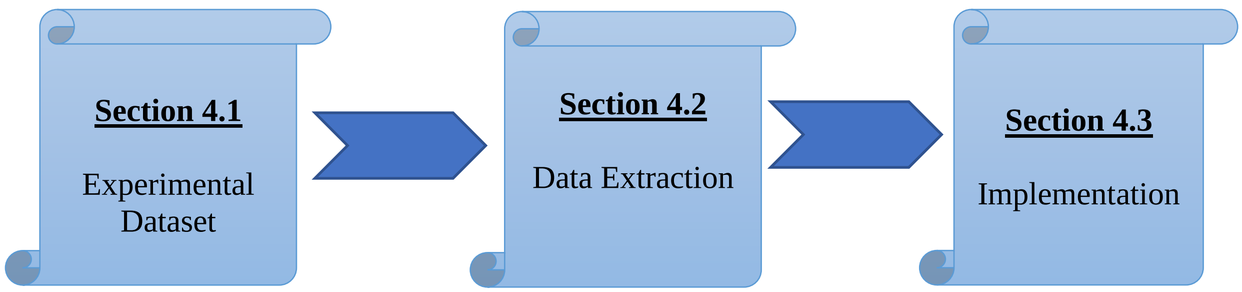
The final results are then compared against the existing results obtained from the existing approaches. This comparison shows how this approach is better as compared to the previous ones.

Chapter 4

Implementation

CHAPTER 4: IMPLEMENTATION

This chapter presents the implementation details for inferring kind and strength of dyadic relationships through detected Wi-Fi access points and Bluetooth devices. Microsoft excel is used to extract data of each individual separately and proposed methodology is implemented on Python. **Section 4.1** described dataset used for analysis. **Section 4.2** explained data extraction and attributes used in this research and **Section 4.3** explained the implementation details of methodology.



4.1 Experimental Dataset

As explained in the previous section, we have used four datasets for the experimental setup. These datasets are named as Movie Review, Customer Review, SST-1, and SST-2. All of the four datasets have been used in previous studies. Preprocessing is performed on each of these datasets.

3.3.5 Stanford Sentiment Treebank

In this section, we are going to discuss some details about the Stanford Sentiment Treebank. The Sentiment Treebank contains labels for each phrase in thousands of sentences. It helps in the training and evaluation of the results. It has been derived from the movie review excerpts that were available on the site rottentomatoes.com. The original dataset has thousands of sentences with labels Very negative, Negative, Neutral, Positive, Very Positive.

The neutral reviews were removed from SST-1 dataset to transform it into SST.

4.1.2 Preprocessing Steps

We have performed the following Pre-processing steps on the datasets:

- **Tokenization**

Th process of breaking long sentences into small pieces is known as tokenization. The small pieces are known as tokens. Unigrams are those tokens that only have one word. Moreover, those tokens that are comprised of two words are known as bigrams.

- **Stop words removal**

The stop words are basically those sentences that are used commonly in order to complete the sentence. These words are not valuable for sentiment analysis. Some examples of stop words are above, about, the, and etc. All of these words are removed so that the processing time can be improved.

- **Special characters removal**

This category of characters includes all those that are not called alphanumeric. We have extracted the words that are necessary for the sentence polarity by removing special characters.

- **Abbreviations Removal**

Most of the reviews that are shared online contain abbreviations. We all know that abbreviations have no contribution to the polarity of the review. So all the abbreviations were removed in the cleaning phase.

4.2 Implementation of RNN-LSTM methodology

There are many deep learning models that have been used in the previous researches. Some of these examples are CNN, RNN, LSTM, RNTN, etc. We have performed many experiments to check and see which one of them is performing better as compared to others. The results showed that RNN and the LSTM model provide the best results for sentiment classification. These models achieved the best F1-score as compared to all the other deep learning models. The results of the experimentation proved that RNN-LSTM model out performs in terms of accuracy and F1-score together as compared to when they are used individually.

4.3 Baseline Methods

The following baseline methods have been used in the previous studies for sentence-based sentiment classification.

- **MNB**
  Multinomial Naive Bayes model that is implemented with unigrams
- **NBSVM**
  It is a variant of SVM that uses the feature values of log-count ratios obtained from naive Bayes [28].
- **Tree-CRF**
  It is basically a sentiment classification method based on the Dependency tree. It uses a combination of hidden variables with CRF [29].
- **RAE**
  These are recursive autoencoders that are semi-supervised [30].
- **MV-RNN**
  It is basically a Recursive neural network. It performs semantic compositionality by using a matrix and a vector on every node that exists in the parse tree [31].
- **RNTN**
  It is a Recursive deep neural network that uses a tensor-based feature function. This model is used for semantic compositionality on a sentiment dataset [32].

- **Paragraph-Vec**

  It is an unsupervised algorithm used for sentence and document classification. It uses them to learn distributed feature representations [33].

- **DCNN**

  It is a Dynamic convolutional neural network used for sentiment classification. It uses k-max pooling operation to perform the tasks [34].

- **CNN-non-static**

  It is a 1-dimensional Convolutional Neural Network. This model has a finetuning optimizing strategy and uses a pre-trained word embeddings [35].

- **CNN-multichannel**

  It is also a 1-dimensional Convolutional Neural Network that uses pre-trained word embeddings in two sets [35].

- **DRNN**

  This is basically a Deep recursive neural network that contains multiple recursive layers stacked over each other [36].

- **Multi-task LSTM**

  It uses a multi-task learning framework that is used for classification purposes. In this model, the learning across various related tasks is done by using LSTM [37].

- **Sentic patterns**

  This sentiment analysis technique uses dependency-based rules for the prediction of sentiments [38].

- **BiLSTM-CRF-CNN**

  This technique uses BiLSTM-CRF to classify opinionated sentences into three types and feeds each group to one-dimensional CNN for sentiment classification. [10]

- **Tree LSTM**

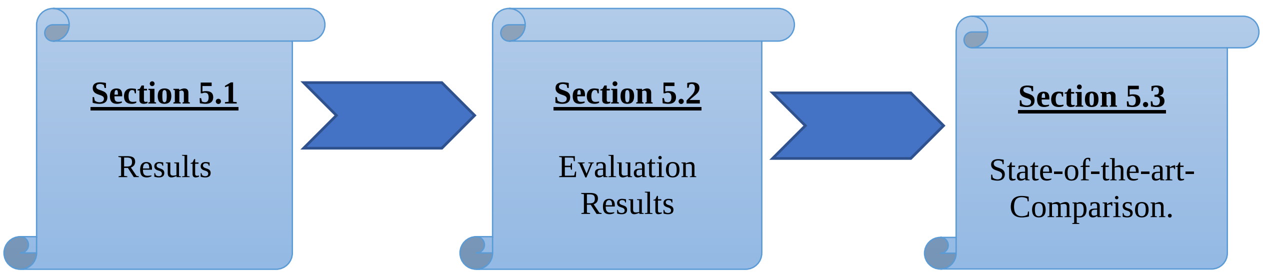
  This approach presents the network topologies in the form of a tree structure [11].

Chapter 5

Results and Discussions

CHAPTER 5: RESULTS AND DISCUSSIONS

This chapter presents results that were obtained using the proposed methodology and comparison with other studies. **Section 5.1** explained the results and **Section 5.2** presents evaluation results and **Section 5.3** presents state-of-the-art- comparison.



5.1 Evaluation Results

We have used two performance measures in order to verify the results of the proposed technique. These measures are accuracy and the F1 score. We have mentioned the evaluation results obtained from all the four datasets in **Table 6.**

Figure 4 represents a comparison between MR, SST-1, SST-2 and CR datasets based on accuracy and F1-Score.

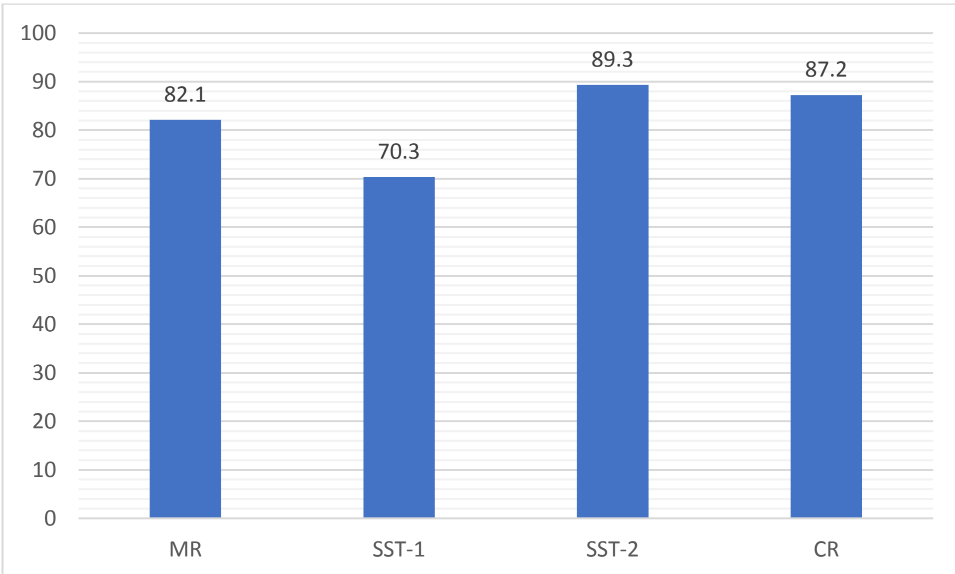


Figure 4: Comparison Between Datasets

5.2.1 Accuracy

Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

5.2.2 F1-Score

F1-Score is calculated as:

$$\text{F1-Score} = \frac{2^{*}\,(\text{True Positive})}{2^{*}\,(\text{True Positive}) + \text{False Positive} + \text{False Negative}}$$

- TP are the number of the key Players that are identified correctly by the model known as true positive
- TN is called the true negatives. It is the number of the normal nodes that are identified correctly.
- FP are false positives, it defines the number of the normal nodes that are identified as the key Players.
- FN also called False Negatives, it illustrates the number of the key Players that are identified as the normal during the classification phase.

Classification is performed using all the four datasets using our proposed model RNN-LSTM. You can see that the model performs best on SST-2 dataset.

Table 4: Classification Results of RNN-LSTM Model

Sr#	**Dataset**	**Accuracy**	**Precision**	**Recall**	**F-1 score**
1	MR	82.1	0.82	0.82	0.82
2	SST-1	70.3	0.70	0.70	0.70
3	**SST-2**	**89.3**	**0.89**	**0.89**	**0.89**
4	CR	87.2	0.87	0.87	0.87

5.2 Comparison

We have conducted this research in order to overcome the research gaps that existed in previous studies. There are thousands of researches that have been conducted in the past. All of them proposed different models and techniques for sentiment classification.

The technique that we have proposed is based on the best deep learning models. We have done a fair comparative analysis by selecting the techniques used in the past and tested them on all the four datasets. Our results are compared with the best models used in the past researches.

Table 7 shows that our proposed model RNN-LSTM performs better as compared to other techniques.

Table 7. Comparison Based on Techniques

Model	MR	SST-1	SST-2	CR
MNB	79.0	-	-	80.0
NBSVM	79.4	-	-	81.8
Tree-CRF	77.3	-	-	81.4
Sentic patterns	-	-	86.2	-
RAE	77.7	43.2	82.4	-
MV-RNN	79.0	44.4	82.9	-

RNTN	-	45.7	85.4	-
Paragraph-Vec	-	48.7	87.8	-
DCNN	-	48.5	86.8	-
CNN-non-static	81.5	48.0	87.2	84.3
CNN-multichannel	81.1	47.4	88.1	85
DRNN	-	49.8	86.6	-
Multi-task LSTM	-	49.8	86.6	-
Tree LSTM	-	50.6	86.9	-
BiLSTM-CRF	82.3	48.5	88.3	85.4
Our approach RNN-LSTM	82.1	**70.3**	**89.3**	**87.2**

The above table shows that RNN-LSTM outperforms all other classification techniques. The second technique that performs better is BiLSTM-CRF that has been used as a baseline in this research. The next one is CNN-multichannel that provides the 88.1 accuracies. So all of these approaches can be used for comparison purposes.

Table 8: Comparison Results of Different Sequence Models

Model	MR	SST-1	SST-2	CR
CRF	81.7	47.6	87.4	84.4
LSTM	81.3	47.5	87.6	84.1
BiRNN	81.7	48.1	87.9	84.8
BiRNN-CRF	81.8	48.3	87.9	84.9
BiLSTM	82.0	48.3	88.0	85.3
BiLSTM-CRF	82.3	48.5	88.3	85.4
RNN-LSTM	82.1	**70.3**	**89.3**	**87.2**

Table 8 presents the comparison of existing sequence models and the proposed technique. We observe that the accuracy has been improved when we have used the new model. Hence we can say that RNN-LSTM is the best performing classifier.

Figure 5. Performance Measures Based on Techniques

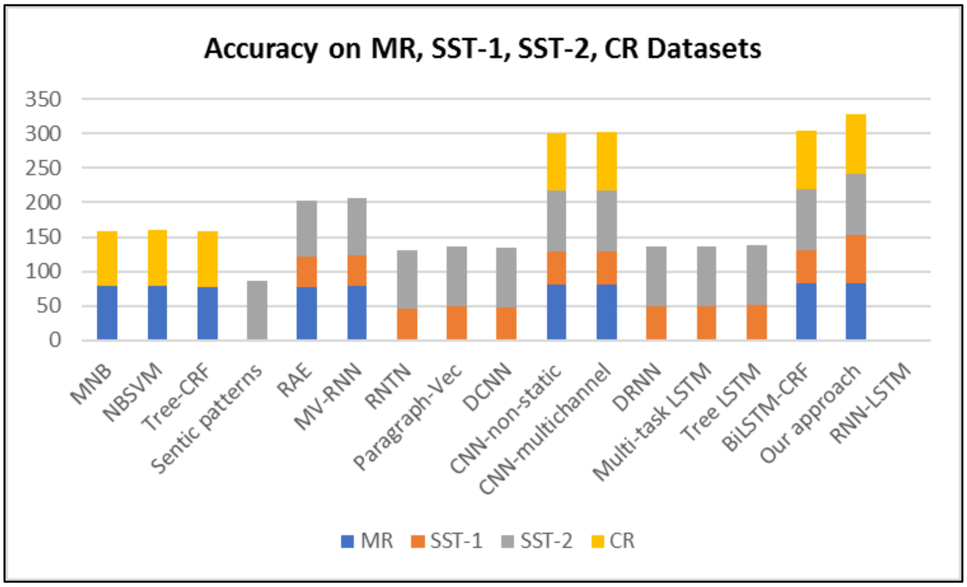


Figure 6. Performance Measures Based on Different Sequence Models

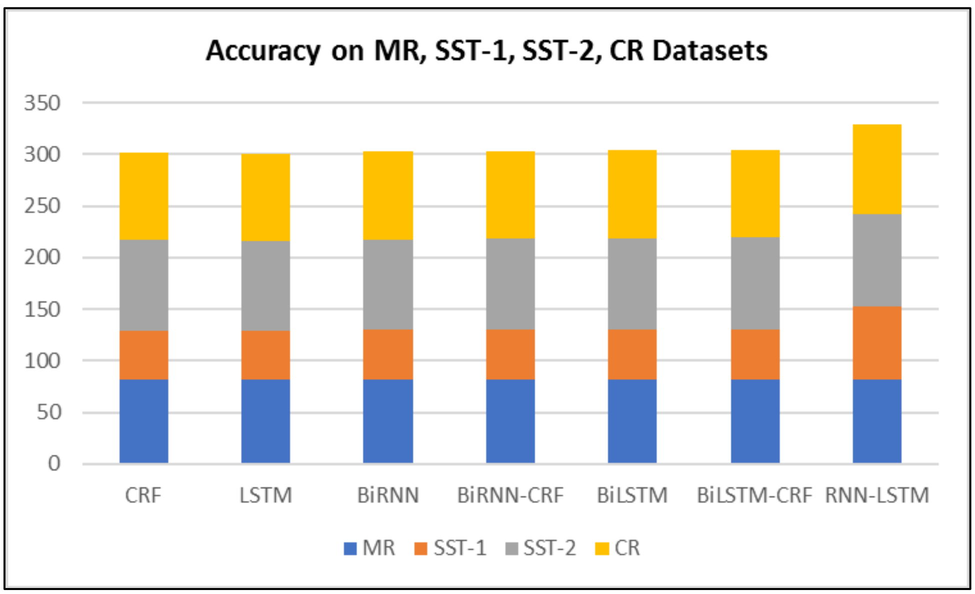
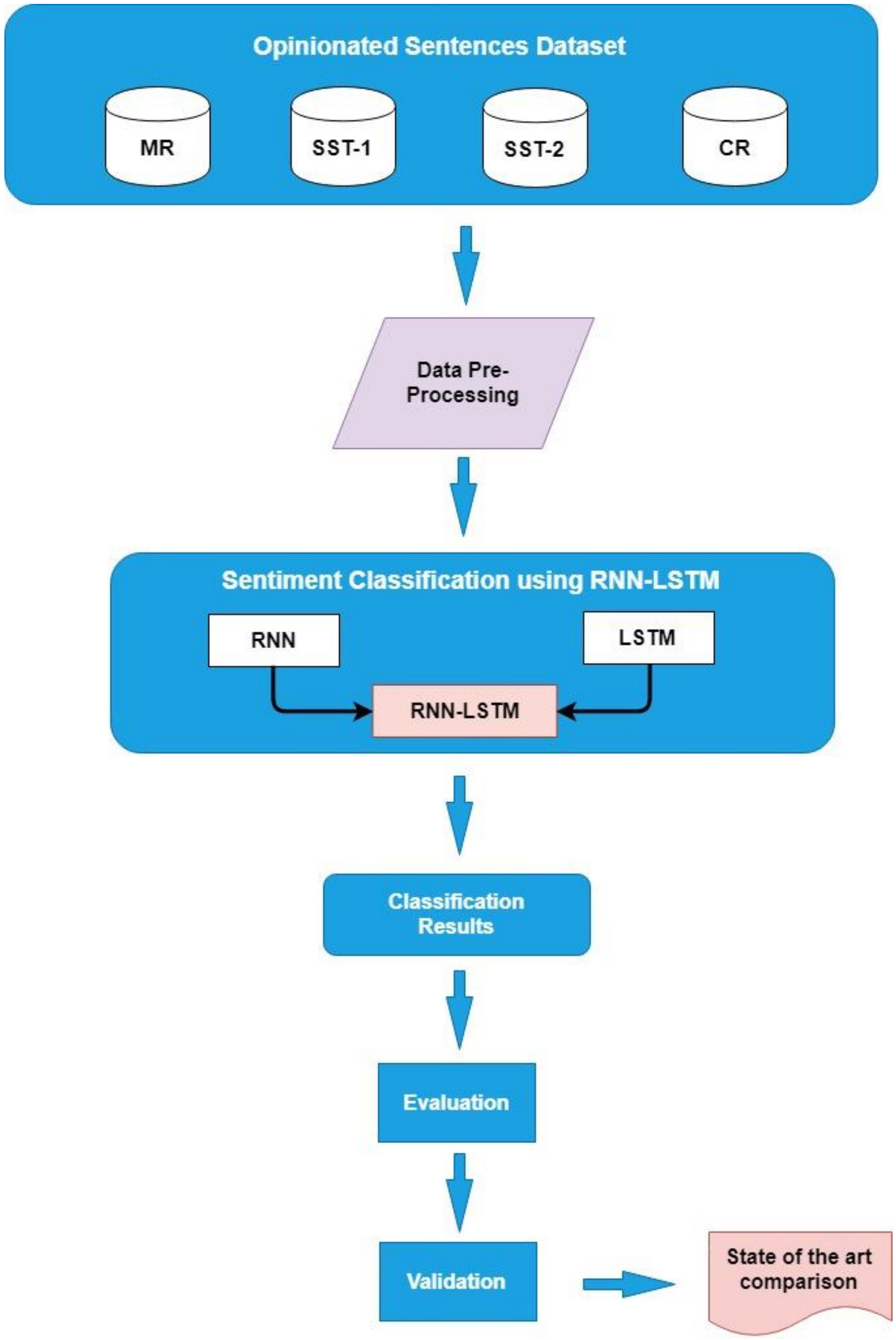


Figure 7. Best Performing Model (RNN-LSTM)



5.3 Discussion

Deep Learning has emerged as a new domain that is being explored by researchers who are interested in working in the field of machine learning. This technique has proven advantages over traditional machines learning models. This technique outperforms all other approaches in most of the machines learning tasks. Researchers who have applied the deep learning technique for Natural Language Processing have achieved a proven success in the past few years. The results that have been achieved from deep learning models for performing Sentiment Analysis tasks are impressive. However, there are still some short comings that need to be addressed. Researchers need to investigate the existing NLP and deep learning techniques to achieve better results for sentiment analysis tasks. This is a major motivation that forced us to conduct this research.

The purpose of this research study is to evaluate effectiveness of all the deep learning models and then to overcome the major shortcomings that exist in these approaches. We propose a deep learning based RNN-LSTM technique that can be used for performing Sentiment Analysis tasks on opinionated sentences. It is evident from the results that three CNN based methods perform the best results on CR and MR data sets. We can understand the effectiveness of Deep learning techniques from this finding. Furthermore, our approach has provided the best results on three out of four datasets as compared to other baseline models. For example, our approach gives an improved accuracy of 70.3 %, 89.3% and 87.2 % on SST-1, SST-2 and CR datasets respectively. However, the BiLSTM-CRF approach achieved an accuracy of 48.5%, 88.3% and 85.4% respectively.

In short, the major contributions of this research study are to propose an efficient Deep Learning model to predict sentiments with improved accuracy. We have tested the proposed model on four bench mark datasets. We have also compared the results with existing machine learning models to highlight the contribution of our newly proposed model. The results proved that it is now possible to predict Sentiments with improved accuracy from opinions of varying length with improved accuracy.

5.4 Limitation

We have predicted a model that produces state of the art results on opinionated sentences obtained from four English language datasets. Our results indicated that the RNN-LSTM model performs best for sentiment analysis in that case. However, there are a few limitations to our work. We have not extended the experimentation on other popular datasets available online. It is a work in progress and we plan to present the results very soon.

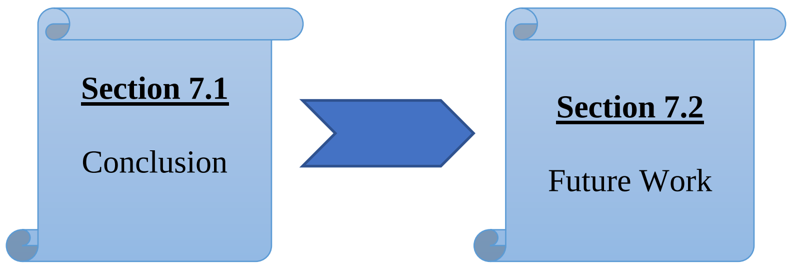
Moreover, at times deep learning models perform differently on different languages. We have only tested our approach to English language datasets. We have not tested our approach on other languages such as Spanish, French, Russian, Dutch, Turkish and others. Hence, it would be interesting to see how the proposed RNN-LSTM model performs on other languages as well.

Chapter 6

Conclusion and Future work

CHAPTER 6: CONCLUSION AND FUTURE WORK

This section presents the Conclusion (**Section 6.1**) and Future Work **(Section 6.2)** of research work.



6.1 Conclusion

There are hundreds of models available that can be used for predicting sentiments from opinionated sentences. Some of them have proved to be providing better results as compared to others. Sentiment analysis is important because it can help us to make decisions about the opinions of people regarding different products and services. Therefore, it was necessary to identify the gaps in the existing researches and present an accurate model that can provide better results as compared to others. There are many deep learning models that can be used for sentiment analysis. However, the previous researches have just focused majorly on CNN. It was necessary to predict an efficient model that is able to overcome the performance problems in the past researches.

This research study presents a novel approach to improve sentiment analysis for opinionated sentences. The approach uses RNN-LSTM to classify the sentiments based on opinionated sentence dataset. The opinioned sentences are used to train the RNN-LSTM model for sentiment classification. We have done extensive experimentation on four sentence-level sentiment analysis datasets. The results are then compared with 11 different approaches used in the past. The results as a result of the experimentation prove that our model out performs all other approaches on three out of the four datasets. The empirical results show that the performance of sentence level sentiment analysis can be improved by using state of the art deep learning models.

6.2 Future Work

In future, we plan to extend this research to improve its performance for sentence-level sentiment analysis. This research study can be further extended to explore other deep learning models. We have only considered four sentence datasets and extensive experimentation is required to see how the perform on other datasets. This approach can then be used to predict sentiments for product reviews and political sentiments of the general public during elections.

The proposed framework is further under investigation for datasets obtained from different languages such as Spanish, French, Russian, Dutch, and Turkish.

REFERENCES

1 Liu, Bing. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

2 Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.

3 Pang, Bo, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004.

4 Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012): 1-167.

5 Narayanan, Ramanathan, Bing Liu, and Alok Choudhary. "Sentiment analysis of conditional sentences." Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1. Association for Computational Linguistics, 2009.

6 Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

7 Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition." IEEE Signal processing magazine 29 (2012).

8 Sun, Xiao, Chengcheng Li, and Fuji Ren. "Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features." Neurocomputing 210 (2016): 227-236.

9 Rong, Wenge, et al. "Structural information aware deep semi-supervised recurrent neural network for sentiment analysis." Frontiers of Computer Science 9.2 (2015): 171-184.

10 Chen, Tao, et al. "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN." Expert Systems with Applications 72 (2017): 221-230.

11 Poria, Soujanya, Erik Cambria, and Alexander Gelbukh. "Aspect extraction for opinion mining with a deep convolutional neural network." Knowledge-Based Systems 108 (2016): 42-49.

12 Zhou, Shusen, Qingcai Chen, and Xiaolong Wang. "Fuzzy deep belief networks for semi-supervised sentiment classification." Neurocomputing 131 (2014): 312-322.

13 Araque, Oscar, et al. "Enhancing deep learning sentiment analysis with ensemble techniques in social applications." Expert Systems with Applications 77 (2017): 236-246.

14 Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic Tweets using deep learning." Procedia Computer Science 142 (2018): 114-122.

15 Ray, Paramita, and Amlan Chakrabarti. "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis." Applied Computing and Informatics (2019).

16 Jangid, Hitkul, et al. "Aspect-based financial sentiment analysis using deep learning." Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee, 2018.

17 Hasan, Ali, et al. "Machine learning-based sentiment analysis for twitter accounts." Mathematical and Computational Applications 23.1 (2018): 11.

18 Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Improving deep neural network design with new text data representations." Journal of Big Data 4.1 (2017): 7.

19 Parlar, Tuba, Selma Ayşe Özel, and Fei Song. "QER: a new feature selection method for sentiment analysis." Human-centric Computing and Information Sciences 8.1 (2018): 10.

20 Vo, Khuong, et al. "Handling negative mentions on social media channels using deep learning." Journal of Information and Telecommunication (2019): 1-23.

21 Rong, Wenge, et al. "Structural information aware deep semi-supervised recurrent neural network for sentiment analysis." Frontiers of Computer Science 9.2 (2015): 171-184.

22 Poria, Soujanya, et al. "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis." Neurocomputing 261 (2017): 217-230.

23 Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.

24 Chachra, Anmol, Pulkit Mehndiratta, and Mohit Gupta. "Sentiment analysis of text using deep convolution neural networks." 2017 Tenth International Conference on Contemporary Computing (IC3). IEEE, 2017.

25 Wang, Sida, and Christopher D. Manning. "Baselines and bigrams: Simple, good sentiment and topic classification." Proceedings of the 50th annual meeting of the association for computational linguistics: Short papers-volume 2. Association for Computational Linguistics, 2012.

26 Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi. "Dependency tree-based sentiment classification using CRFs with hidden variables." Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

27 Socher, Richard, et al. "Parsing natural scenes and natural language with recursive neural networks." Proceedings of the 28th international conference on machine learning (ICML-11). 2011.

28 Socher, Richard, et al. "Semantic compositionality through recursive matrix-vector spaces." Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Association for Computational Linguistics, 2012.

29 Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.

30 Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." International conference on machine learning. 2014.

31 Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences." arXiv preprint arXiv:1404.2188 (2014).

32 Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).

33 Irsoy, Ozan, and Claire Cardie. "Deep recursive neural networks for compositionality in language." Advances in neural information processing systems. 2014.

34 Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks." arXiv preprint arXiv:1503.00075 (2015).