Capsule layer and attention layer augmentation analysis over convolutional methods for review categorization in Amazon dataset

MAJOR PROJECT REPORT

Submitted in partial fulfillment of the requirements for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

by

Shivani Aggarwal	Navya Mahajan	Abha Tripathi	Sharat Sachin
(35611502716)	(35411502716)	(41211502716)	(03111502716)

Guided By
Ms. Preeti Nagrath
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING
(AFFILIATED TO GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY, DELHI)
DELHI – 110063

APRIL 2020

CANDIDATE'S DECLARATION

It is hereby certified that the work which is being presented in the B. Tech Major Project Report entitled "**Capsule layer and attention layer augmentation analysis over convolutional methods for review categorization in Amazon dataset"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** and submitted in the **Department of Computer Science & Engineering** of **BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING, New Delhi (Affiliated to Guru Gobind Singh Indraprastha University, Delhi)** is an authentic record of our own work carried out during the period from **January 2020** to **April 2020** under the guidance of **Ms. Preeti Nagrath, Assistant Professor.**

The matter presented in the B. Tech Major Project Report has not been submitted by me for the award of any other degree of this or any other Institute.

Shivani Aggarwal	**Navya Mahajan**	**Abha Tripathi**	**Sharat Sachin**
(35611502716)	**(35411502716)**	**(41211502716)**	**(03111502716)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge. They are permitted to appear in the External Major Project Examination.

Ms. Preeti Nagrath
Assistant Professor

Prof. Kirti Gupta
HOD, CSE

The B. Tech Minor Project Viva-Voce Examination of Sharat Sachin (03111502716), Abha Tripathi (41211502716), Shivani Aggarwal (35611502716), Navya Mahajan (35411502716), has been held on …….

(Signature of External Examiner)

Dr. Pranav Dass
Project Coordinator

Ms. Nikita Jain
Project Coordinator

ABSTRACT

Text sentiment analysis is an important and challenging task, and sentiment analysis of customer reviews is a common problem faced by companies. The world as we know it today is becoming more and more technologically advanced. E-commerce is one of the most widely adopted of these advancements, now seeing penetration all around the world. Now people sitting in the comfort of their homes can order items and have them at their fingertips in days, and some cases, hours. For these people relying on the online items, one of the most important metrics is the reviews received by that item. To obtain a balanced view of the item, the customer may need to see the positive as well as negative reviews for that item. Modern day algorithms and models can be made to decide the reviews seen by the user so he can judge an item fairly. We have attempted to make models that can judge the sentiment of the user writing the reviews from the text of the review, classifying it into positive, negative or neutral. In this work, we have performed a comparison of the effect of convolutional layers, attention layers and capsule network layers on base models for GRU, Bi-GRU, LSTM and Bi-LSTM on an Amazon review dataset used for classification.

ACKNOWLEDGEMENT

We express our deep gratitude to **Ms. Preeti Nagrath**, Assistant Professor, Department of Computer Science & Engineering for her valuable guidance and suggestions throughout our project work. We are thankful to **Dr. Pranav Dass** and **Ms. Nikita Jain**, Project Coordinators, for their valuable guidance.

We would like to extend our sincere thanks to Head of the Department, **Prof. Kirti Gupta** for her time to time suggestions to complete our project work. We are also thankful to **Prof. Dharmender Saini**, Principal for providing us the facilities to carry out our project work.

Shivani Aggarwal (35611502716)	**Navya Mahajan (35411502716)**	**Abha Tripathi (41211502716)**	**Sharat Sachin (03111502716)**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

RNN	Recurrent Neural Network
DNN	Deep Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Time Memory
Bi-LSTM	Bidirectional Long Short-Time Memory
Bi-GRU	Bidirectional Gated Recurrent Unit
Bi-RNN	Bidirectional Recurrent Neural Network
NLP	Natural Language Processing
G RNN	Gated Recurrent Neural Network
CNN	Convolutional Neural Network
SVM	Support Vector Machine
TGRU	Two Stream Gated Recurrent Unit
SSWE	Sentiment Specific Word Embedding
POS	Part of Speech
ABSA	Aspect Based Sentiment Analysis
LDA	Latent Dirichlet Allocation
GloVe	Global Vectors for Word Representation
NB	Naïve Bayes
RNTN	Recursive Neural Tensor Network
CRF	Conditional Random Field

LIST OF SYMBOLS

c_t^i	j^{th} LSTM unit preserves a memory
h_t^j	output or the activation of the LSTM unit
t	time
o_t^j	output gate that controls the amount of exposure to the memory content
V_0	diagonal matrix
σ	logistic sigmoid function
c_t^j	memory cell
$\underline{c}_t^j$	new memory content
f_t^j	forget gate
i_t^j	input gate
V_f & V_i	diagonal matrices
h_t^j	activation
z_t^j	update gate
$\underline{h}_t^j$	candidate activation
r_t^j	reset gate
$\odot$	element-wise multiplication
r_t	calculated set of reset gates

CHAPTER 1 : INTRODUCTION

Sentiment analysis, also known as opinion mining, is the art of extracting emotion from blocks of text, with no help from human actors so as to categorize them according to the polarity of the sentiment. We attempt to glean positive or negative intent from text, and try to also categorize cases where neither of the positive and negative overwhelms the other and the overall intent can thus be classified as neutral. It is a major research area in natural language processing, data mining, and machine linguistics. Examining and processing the full text, sentiment analysis tries to extract sentiment information from it.

With the rate at which internet usage has changed for people in recent times, the different manners in which production and dissemination of content occurs become more and more diverse, especially in terms of the different forms of content generated by the users [1]. A significant part of this contains phrases from which humans can glean emotional awareness, yet this poses significant challenges for a learning system. A business can gain a significant advantage by keeping abreast of the consensus relating to its products online, on interfaces such as review sites and product forums. This can also influence decisions made concerning customers [2]. It helps companies process and extract precious data of high value to improve brand image, and focus their attention on improving aspects of the business that have low customer satisfaction [3].

RNN's are models that can be applied to a variety of problems dealing with sequenced data, however they are limited by the exploding and vanishing gradient problems [4]. The vanishing gradient problem was successfully resolved by Hochreiter & Schmidhuber [5], with the development of the LSTM architecture, which is insusceptible to this particular issue. The GRU (introduced by Kyunghyun Cho et al in 2014 [6]) performs all the functions performed by the LSTM unit, but it accomplishes these functions without employing any extra memory units. The bidirectional variants of these models are designed with a specific motivation in mind - that the context of a word is defined by what precedes that word as well as what follows it. Thus, these models act with data from both backwards and forwards in time to make a decision. The capsule neural network (CapsNet) is a system that attempts to add structures that better mimic relationships that contain a hierarchical collection of objects. A capsule is a group of neurons that activate uniquely for various characteristics of a different object, and we reuse output from these capsules to form more balanced representations for higher ranked capsules [7].

Deep learning-based models have shown great promise in determining the polarity of text in recent times, with a variety of research focusing on text classification by polarity [8]. One reason for this is that they are able to learn the intent of the writer from training data without tiresome feature engineering. Sentiment analysis is a very important field of natural language processing and has been studied in various fields, including news, social media, movies and product reviews [9]. In this paper, we are comparing the results of using baseline unidirectional and bidirectional LSTM and GRU to ones augmented using a convolutional layer and self-attention layer. We also use the state-of-the-art capsule network layer to augment the base models and compare the accuracy achieved with them.

Descriptions of the convolutional neural networks, gated recurrent architectures, attention mechanism and capsule network are discussed in the second chapter in this report. The next chapter consists of the literature review, that consists of descriptions of the previous research performed in various fields of deep learning, namely text classification based on aspect sentiment analysis, baseline deep learning methods, capsule networks and attention mechanism. Details about the methodology used, proposed system and evaluation measures are mentioned in the fourth chapter, and in the fifth chapter all results and analysis of the research performed are included. Finally, we give a conclusion in the sixth chapter.

CHAPTER 2 : DEFINITIONS

2.1 Convolutional Neural Networks

A Convolutional neural network (CNN) is categorized as a deep, feed-forward artificial neural network that is generally used in fields like pattern recognition; from image processing to voice recognition [10] or in classification, segmentation. In this model, it encompasses neurons, represented by three dimensions, the spatial dimensionality of the input (height and the width) and the depth, which are self-optimizing in nature. It is a type of deep learning architecture that is usually composed of a set of layers namely, convolutional layer, pooling layer, and fully connected layers. Each of these layers has unique parameters that can be used for optimization and will perform different tasks on input data.

1. Input layer - It will hold the input data of the dataset.
2. Convolutional layer - Convolutional layers, related to feature extraction are the type of layers where filters are applied and it comprises parameters such as a number of 'kernels' or filters and size of 'kernels' or filters. It is a type of layer where all the user-defined parameters are present and consist of filters called 'kernels.
3. Pooling layer- It consists of performing the process of extracting a particular value from a set of values, specific operations are performed such as max pooling, average pooling, and min pooling to basically reduce the dimensionality of the network.
4. Fully connected layer- This layer forms the last block of the CNN architecture, related to the task of classification. It is a type of layer that takes input from the previous layer for the classification output of a CNN and is used to provide more flattened results.

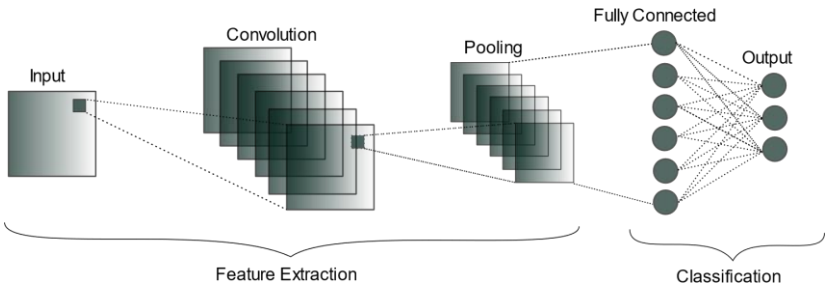

Figure 2.1 Framework for convolutional neural network [20]

2.2 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) have shown great potential in machine translation tasks [13] and also is a type of artificial neural network that can be used in applications like handwriting recognition [14] or speech recognition [15]. Unlike feedforward neural networks, these types of neural networks have inputs connected to each other which will result in temporal behavior for a time sequence which makes these networks more agreeable for sentiment analysis. These RNN are a network with nodes in the form of neurons and follow a particular sequence and which will result in a directed graph, with each node having input from backward nodes and output to forward nodes. Also, these nodes have a real valued weight attached to them which will result in making modifications to the output and the signals passing through them. Given a sentence S, we can obtain hidden states $H = [h_1, h_2, \ldots, h_n]$ by

feeding the input $X = [x_1, x_2, x_3 \ldots, x_n]$ through RNN with dim_w amd dim_h, dimensions of word embedding and the hidden states respectively.

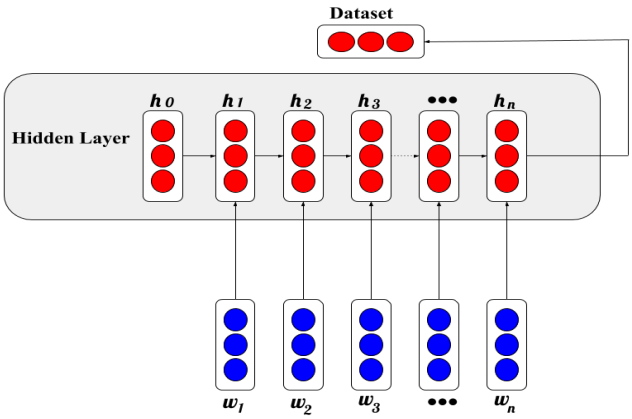

Figure 2.2 Framework for recurrent neural network

Although RNN's have several shortcomings which will make them difficult to model and train such as vanishing gradient problem or exploding gradient problem or it becomes difficult to process very long sequences if we are using 'relu' or 'softmax' function as an activation function. Both infinite and finite impulse type RNN may consist of an additional stored state, however, this storage can directly be controlled by the neural network. Therefore, the storage may also be replaced by another graph or network which has integrated time delays or feedback loops. The most effective solution to this problem is adding a gating mechanism to the RNN. The framework for an RNN is described in figure 2.2.

Two gated RNN used are:

- Long Short-Term Memory (LSTM) [5]
- Gated Recurrent Unit (GRU) [6]

Look at figures 3(a) and 3(b) respectively for illustrations of the LSTM unit and GRU units respectively.

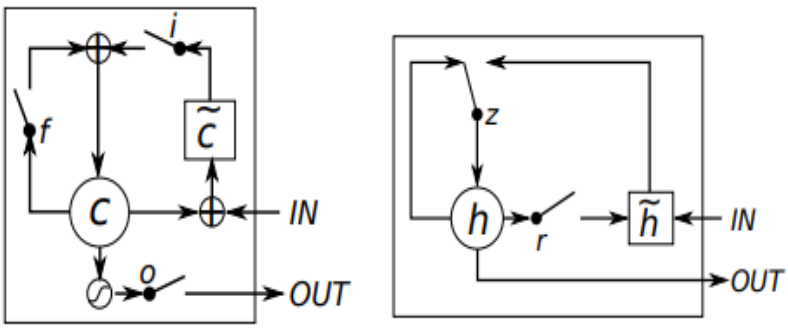

Figure 2.3 **(a)** LSTM Unit [19] **(b)** GRU [19]

2.3 Long Short-Term Memory (LSTM)

The LSTM [5] unit was developed by Hochreiter and Schmidhuber. It has since undergone many changes over the years like the addition of forget gate. LSTM consists of a memory cell for the storage

of information. It computes the input, output and forget gate to manage this memory. LSTM units can thus perform delicate tasks like propagating or keeping the flow of an important feature that came early compared to others in the input sequence worked over a long distance. LSTM despite being complex is very successful in various tasks like handwriting recognition, machine translation and of course sentiment analysis.

Unlike the vanilla recurrent unit which calculates a weighted sum of all the input signals and also at time t applies a function (nonlinear) each j^{th} LSTM unit preserves a memory c_t^i. The output h_t^j [11], or the activation of the LSTM unit, is then,

$$h_t^j = o_t^j \ tanh\ tanh\ (c_t^j) \tag{1}$$

where o_t^j is the output gate that controls the amount of exposure to the memory content.

The output gate is then computed as,

$$o_t^j = \sigma(W_0 x_t + U_0 h_{t-1} + V_0 c_t)^j \tag{2}$$

where V_0 is a diagonal matrix and σ is a logistic sigmoid function.

The memory cell c_t^j is updated by forgetting the existing memory to a limited extent and adding new memory content $\underline{c_t^j}$,

$$c_t^j = f_t^j\, c_{t-1}^j + i_t^j\, \underline{c_t^j} \tag{3}$$

where the new memory content is defined as,

$$\underline{c_t^j} = tanh\ (W_c x_t + U_c h_{t-1})^{\,j} \tag{4}$$

The degree to which the existing memory is forgotten is modulated by a forget gate f_t^j and the extent to which the memory cell is augmented by the new memory content is modulated by an input gate i_t^j. Gates are computed by,

$$f_t^j = \sigma(W_f x_t + U_f h_{t-1} + V_f c_t)^j \tag{5}$$

$$i_t^j = \sigma(W_i x_t + U_i h_{t-1} + V_i c_t)^j \tag{6}$$

Note that V_f and V_i are diagonal matrices.

2.4 Gated Recurrent Unit (GRU)

A slightly more sudden variation on the LSTM is the Gated Recurrent Unit, or GRU, introduced by Cho, et al [6]. It only has two gates, a reset gate and an update gate (a combination of the forget and input gates). The resulting model is simpler and easier to compute as compared to standard LSTM models. GRU's got rid of the cell state that means it doesn't have any extra separate defined memory cells and thus uses the hidden state to transfer information.

At a particular time *t* of the GRU, the activation h_t^j is defined as a linear interposition between the previous activation h_{t-1}^j and the new candidate activation [11] h_t^j,

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \underline{h_t^j} \tag{7}$$

where an update gate resembled by z_{ret}^j is responsible to tell that how much the unit changes its content or activation. The update gate is calculated as,

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j \tag{8}$$

The candidate activation $\underline{h_t^j}$ is calculated exactly in the same manner to the traditional recurrent unit as,

$$\underline{h_t^j} = tanh\,(Wx_t + U(r_t \odot h_{t-1}) \quad (9)$$

where $\odot$ represents element-wise multiplication and r_t is defined as a calculated set of reset gates.

When turning to off (r_t^j tending to 0), the reset gate is responsible that makes the unit behave in a manner as if it is extracting to read the very first symbol of the given input sequence, and further allows it to remove the already existing calculated state. The reset gate r_t^j is calculated in the same manner to the update gate,

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \quad (10)$$

2.5 Bidirectional-Gated Recurrent Unit (Bi-GRU)

Interest in incorporating a Bi-GRU [11] layer into the architecture of our segmentation model stems from their ability to combine cell state and hidden state into one which will result in faster training, particularly when training on large datasets.

The Bi-GRU layer has two parts basically which simultaneously read the word vector from the forward and reverse directions to make predictions about the current state. Bi-GRU's are a type of bidirectional RNN, deep learning architecture with only forget and input gates.

2.6 Bidirectional-Long Short-Term Memory (Bi-LSTM)

As an extension to the traditional LSTM layer, Bi-LSTM layer [12] is an RNN, generative form of deep learning that came into existence in 1997 by Schuster and Paliwal and uses two LSTMs to learn each token of the sequence, one from right to left and other from left to right. It is widely useful in applications like sentiment classification, sentence classification or handwriting recognition where the learning problem is sequential and there is a need to understand the context more efficiently.

In Bi-LSTM, information or input is not fixed and flows in two directions, one from past to future and other from future to past and thus will result in increasing the amount of input information available to the network. This is achieved using two parallel layers, i.e., a backward hidden layer and forward hidden layer, to generate the output sequence. By taking advantage of two-time directions, input data from the past and future of the current time frame can be used, which will result in fewer delays as compared to unidirectional RNN. Bi-LSTM architecture comes into existence from their ability to recognize long-term dependencies, solve vanishing gradient problems and contextual features from previous and future states which provide them an advantage over the unidirectional.

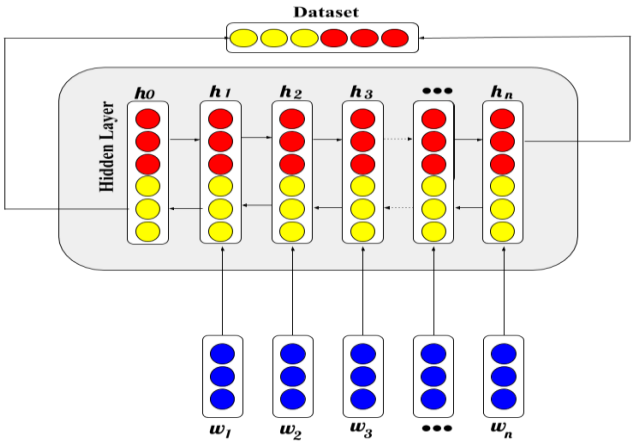

Figure 2.4 Framework for Bidirectional Recurrent Neural Network (Bi-RNN)

2.7 Attention Mechanism

The attention mechanism was introduced by [16] and is widely used for improving various types of neural networks such as CNN, RNN. Self-attention is the process in which we apply attention mechanisms to every position of the source sequence. In this type of attention, we create 3 vectors namely, query, key, value for each sequence position, and then applying the attention for each position x_i, using the x_i query vector, and key and value vectors for all other positions. As a result, an input sequence X of words is transformed into a sequence Y where each y_i contains all the information about x_i .

The above mentioned 3 vectors can be created by applying linear projections which are learned in nature [17], or using feed-forward layers. This computation can be done for the entire source sequence in parallel by grouping the queries, keys, and values in Q, K, V matrices [17].

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{11}$$

2.8 Capsule Network

Capsule network is a type of artificial neural network that was first introduced in 2011 by Geoffrey Hinton [18] as an alternative to CNN. Capsule networks are used to better model hierarchical relationships and also to solve problems caused by max pooling and deep neural networks. It is a type of network that contains numerous capsules which contain groups of neurons which are used to learn or detect an object. These neurons contain all the important or relevant information related to the features such as position, size and hue and are represented in the vector form as an output. To classify different capsules more effectively, a nonlinear function called "squashing" function is introduced into the training procedure, which acts as a part of the activation in the level of the capsule.

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1+\|\boldsymbol{s}_j\|^2}\frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|} \tag{12}$$

where $\boldsymbol{v}_j$ is the vector output of capsule j and $\boldsymbol{s}_j$ is its total input.

A capsule network architecture is organized in multiple layers very much like any neural network. The capsules in the lowest layer are called primary capsules: each of them receives a small region of the image as input, and it tries to detect the presence of a particular pattern, like a rectangle. Capsules in

higher layers, called routing capsules, detect larger and more complex objects. Capsules interact with each other through an iterative “routing-by-agreement” mechanism which was introduced by Geoffrey Hinton [7]. A lower-level capsule prefers to send its output to higher-level capsules. Capsules are very promising in applications such as emotion detection, hand-written and text recognition, machine translation and many more.

CHAPTER 3 : RELATED WORKS

3.1 Text Classification based on aspect-based analysis

To justify the classification of polar sentences, a methodology was implemented based on the aspect-based analysis. Zhuang Chen et al. [21] proposed a model named TransCap (Transfer Capsule Network) basically used to conduct aspect level sentiment classification with document-level knowledge. Also, in this model, they used both the Aspect routing approach for encapsulating the semantic representations and Dynamic Routing approach to adaptively couple the semantic capsules. The Experiment was performed on two SemEval datasets that show the proposed model outperforms other baseline models.

Yunlong Liang et al. [22] worked on the concept of aspect-based sentiment analysis in order to predict and enhance the accuracy of the polar cases of text classification. For this, they worked on Aspect-guided-deep-Transition (AGDT), which consists of aspect-concatenated embedding, aspect-reconstruction, and aspect guided encoder. Further, they used several baseline models like CNN, TD-LSTM, IAN, RAM, and GCAE. They achieved the highest accuracy of 87.72%, in the DS model.

Tao Yang et al. [23] created a new target subnetwork to capture the main fundamental information which consists of conceptual and textual information. In regard to the previous studies where only a polar type of cues was used, these contributed by creating approximately five different types of lexical cues for utilizing the exact information accumulated through words. To utilize the dependency between several cues and words, the authors created a new attention mechanism. Further, they used various baseline models like LSTM, RAM, CueNet, etc. out of which the maximum accuracy obtained was 88.6%.

Yanyan Wang et al. [24] worked on enhancing the previous deep learning technologies and applied a new methodology of SenHint on the English and Chinese benchmark datasets. These datasets were polar in nature due to which they performed sentiment-based analysis and implemented Markov-Logic Network. Finally, the outcome comparison was made in which SenHint-rel was found better than GCAE.

3.2 Analysis of datasets using baseline deep learning models

Various deep learning models are being used for classification purposes. These models in combination with several other baseline models serve features and improvised vector functions to enhance the accuracy and reduce dimensions. These models are also used in mapping sequences of outputs.

Ashraf Elnagar et al. [25] proposed a new single label corpus called the SANAD corpus which contains 200k articles so as to perform text categorization on Arabic news articles. Along with the available categorization systems, several deep learning classifiers were also introduced so as to reduce the need for preprocessing. In this paper, they implemented 9 models namely CNN, RNN, and attention-based models. The results draw a conclusion that DNN models performed best on The SANAD corpus (maximum with HANGRU 95.18% and minimum with CGRU 93.43%).

Tanjim Ul Haque et al. [26] used a supervised learning method on the amazon product reviews dataset to polarize those reviews and get better results. In this model, they use both approaches - active and

manual for the labeling of datasets. In this experiment, they selected three categories- Electronics reviews, Cell Phone, and Accessories Reviews and Musical Instruments product reviews from the amazon dataset along with feature extraction of 3 types: - the bag of words and tf-IDF & Chi-square approach for getting a significant improvement in the accuracy. The results show that 10-fold provided better accuracy than 5-fold and also a linear support vector machine provides the best results with 94.02% accuracy.

Anirban Mukherjee et al. [27] used different types of RNN (simple RNN, GRU, LSTM and Bidirectional LSTM) on the same dataset to find out the model which performs best in the multi-class classification of sentiment analysis. Then, they use that model and apply oversampling to the dataset to show the improvement inaccuracy. In this experiment, they observe that bidirectional LSTM shows the best results without oversampling and after oversampling, the results show a significant increase with 80% accuracy.

Ryan Ong [28] started experimenting by testing out different variations of RNN models including LSTM and GRU both unidirectional and bidirectional with CNN layers. Also, he used different kinds of word embeddings to see if there is a significant change in the results. The results show that Bi-LSTM-CNN shows significant results and therefore they conducted a manual search for some of the key hyperparameters like the level of epochs, spatial dropout rate and Pre-trained vs No pre-trained embeddings. Also, from this experiment conclusion was made that ordering of layers is very important and shows a significant change in the results.

Sharat Sachin et al. [29] conducted a survey for different deep learning models such as LSTM, GRU, Bi-LSTM, and Bi-GRU on the amazon reviews dataset. In this model, they have 3 layers- embedding, unidirectional/bidirectional and dense layers. In this paper, the conclusion was drawn that bidirectional performs better than unidirectional and also GRU's are faster to train than LSTM. Bi-GRU outperforms all other RNN techniques in terms of accuracy.

Gonzalo A. Ruz et al. [30] proposed a methodology to perform sentiment analysis on twitter dataset using Bayesian network classifiers. In this dataset, they classified the sentences into positive type tweets and negative type tweets. Along with the BF-TAN, they deployed a model of SVM with random forest classifiers to compare the performance on Spanish texts as well. To tackle the problem of class imbalance they used SMOTE features. The final result was enhanced by the performance of SVM that increased their accuracy above 80% compared to all the other modules present in the processing part.

3.3 On the basis of Attention mechanism

While implementing sequence modelling networks, the entire details and information regarding input is contained inside the current state of various CNN and LSTM, BI-GRU. There is no doubt that this information is precise and complete. Assumption mechanism deploys the ease of looking forward in the entire hidden state of input and helps in enhancing the computational accuracy. Various researchers have worked upon utilizing the attention mechanism.

Liang Zhou et al. [31] presented a new model Bi-GRU attention-based model to perform text classification on Chinese texts to solve the problem of insufficient text context information learning and weak feature extraction ability. Their proposed model was divided into three parts: the input layer, hidden layer, and output layer. The input layer consisted of the CBOW model in word2vec. The hidden

layer contains BI-GRU. The output layer shows the output of BI-GRU calculated by the softmax function. The accuracy attained by this model was 90.45% and it outperforms the other existing models.

Similarly, Jingren Zhang et al. [32] restructured the formulation of CNN models that are generally used as a deep learning module for text classification. They applied their new proposed mechanism on MRD and SamEval2016 datasets. The new fusion model consists of MATT(CNN) along with the combination of Bi-directional GRU (Bi-GRU). For dimensionality reduction they applied Principle Component Analysis (PCA) This improved their accuracy by 5.94% on the MRD dataset and 11.01% SamEval2016 dataset. The highest accuracy was 79.22% in the laptop category.

Yang Liua et al. [33] introduced an attention gated layer to amplify the output produced by CNN models for textual extraction. According to the analysis and results section, their model enhanced the accuracy of other standard CNN models by 3.1%. Along with this moving forward in their experiment, they applied several activation functions to outperform the existing Relu, sigmoid, softplus, PReLu, LReLu, etc. the proposed functionality that they formulated was Natural Logarithm rescaled Rectified Linear Unit (NLReLU). The final outcome result was 94.5% in the TREC dataset.

Weijiang Li et al. [34] worked in improving the functioning of previous deep learning baseline models for sentiment analysis. This field includes aspect-based terms also and can easily manipulate the outcome of any model. Their methodology provided an improvised version of bidirectional baseline models along with multi-channel features accommodation. They used MF-Bi-LSTM and SA-Bi-LSTM on MR, SST-5, SST-2 datasets. They achieved the highest accuracy of 89%.

Artaches Ambartsoumian et al. [35] worked on a new model based on self-attention networks for sentiment analysis called the SSAN model. They tested the proposed model on 2 layers of stacked configurations. The results show that the SSAN model achieves better accuracy than other neural networks such as CNN or RNN models. Also, the results show that this model was better in capturing long term dependencies than other models.

Gael Letarte et al. [48] proposed a new model called SANets for text classification. The experiment showed that this model achieved better accuracy on sentiment analysis tasks by 2% but no considerable improvement for topic classification. From this model, a conclusion is drawn that self-attention is important for sentiment analysis as it is shown that for each dataset self-attention improves the performance.

Likun Ji et al. [36] proposed a new model based on self-attention mechanism and bidirectional GRU for sentiment analysis. In this paper they introduced multi head self-attention along with position embedding combined with word embedding. They evaluate the proposed model on IMDB dataset which shows 90% accuracy and thus outperforms other models and also the results show that this model is more effective in extracting information for sentiment analysis.

3.4 Capsule Networks for text classification

Various types of important and valuable information is lost while performing the max-pooling stage in convolutional neural networks in the process of text classification. Therefore, keeping this in mind, a new technology of capsule networks was introduced. Capsule networks were basically made to target the lower level features using routing by agreement strategy. Many researchers have compared their

several experiments and implemented capsule networks in their datasets, showing the enhanced accuracy based on the results achieved.

Haftu Wedajo Fentaw et al. [37] along with his team worked on the CapsNets Model for sentence classification. The performance of this model was compared with RNN and CNN models along with baseline models on the five real-world datasets which show significant results. In this paper, they made many observations such as a multi-phase training approach that can be very helpful to increase performance, also CapsNets models are sensitive to the order of the words which means CapsNet models can capture spatial information better. Also, the results show that the proposed network resulted in increased processing speed compared to other baseline models.

Dynamic routing enhancement was the most important key factor that catalyzes the entire functioning of a capsule network. Wei Zhao et al. [38] explored capsule networks for text classifications with the help of dynamic routing. The results show the effectiveness of capsule networks for text classification. In the proposed model they used 4 layers-> n-gram convolutional layer, primary capsule layer, convolutional capsule layer, and fully connected capsule layer. Here they use 2 capsule network architectures (capsule A and capsule B) to integrate these 4 layers in different ways. They conduct the experiment on 6 benchmarks which includes various text classification tasks. Also, they transferred a single label to multi-label text classifications which shows significant improvements in the results over the baseline methods.

Hongxia Yin et al. [39] proposed a Capsule network model for cross-domain sentiment classification along with Identifying Transferable Knowledge (CITK). They also use Bidirectional Encoder Representations from Transformers (BERT) for pre-training. They perform the proposed model on a real-world dataset (with 2 domains electronic and kitchen domain- positive and negative) and the results show that the CITK model outperforms other states of the art methods significantly and produces competitive results.

Along with the approach to use capsule networks, various researchers implemented these networks along with the aspect-based mechanisms and deep learning strategies.

Chi Xu et al. [40] proposed a model named CAPSAR a capsule net model in order to improve the aspect level sentiment analysis. In the proposed model, the capsule was used to denote the sentiment categories and then with the help of sentiment-aspect reconstruction procedure they inject aspect type information into sentiment capsules. The model is performed on 3 real-world benchmarks and shows superiority over the other models.
To enhance the accuracy by using deep learning layers, it helps in accumulating the important features that are compared with training labeled datasets.

Yongping Du et al. [41] proposed a capsule-based hybrid NN with addition to Bi-GRU that is used to obtain the implicit sentiment information and achieve features over long distances. The model is performed on two text datasets and obtained an accuracy of 82.5%. Also, in this model, they used a self-attention mechanism along with CNN, which gives us an advantage of less training time and simple network structure which results in better performance.

Kejun Zhang et al. [42] along with this team proposed a capsule network SC-BiCapsNet for sentiment classification. In the proposed model there are basically 3 parts-> G-Word vector, Bi-channel representation, and capsule modules. This paper used 2 datasets of IMDB and NLPCC 2014 to evaluate

the performance of the proposed model and the results show that the proposed method outperforms other methods on the IMDB dataset with an accuracy of 92.38%.

Jaeyoung Kim et al. [43] compared the performance between traditional deep learning methods and the use of capsules for text classification. They used several datasets of 20news, Reuters, MR (2004), MR (2005), TREC-QA, MPQA. Further, they represented a simple routing method to outperform the computational complexity created by dynamic routing in the capsule networks. The final result achieved was 87% as trained by simple routing. Xian

Zhong et al. [44] worked in analyzing the difference between the traditional deep learning neural network methods and the latest capsule network technology. Based on the incorrect compression mechanisms, the team proposed an IPC-CapsNet algorithm which was combined with SPT-CapsNet in order to improve the computational complexity. The final complexity and accuracy were improved to 85.89%, which was 84.90% earlier.

Deepak Kumar Jain et al. [45] proposed a methodology to check the sincerity of all those sites and communities that provides questions and answers. This project was made to remove irrelevant content and improve the quality of the concerned material. They applied a deep refinement methodology which comprises C-Bi-LSTM. The resultant model outperformed the capacity and accuracy of SVM and Bi-LSTM. The final result produced the highest accuracy of 96.03%.

Min Yang et al. [46] investigated the transferring capability in the field of text classification. Their capsule model outperformed the performance of different models on 4 out of 6 datasets. They used capsule compression and class-guided routing in the dynamic routing mechanism of the final capsule model. They achieved the highest accuracy of 82.3%. Further, they investigated various combinations of baseline models along with capsule networks.

CHAPTER 4 : METHODOLOGY

4.1 Dataset Used

We have used a large dataset of reviews from Amazon developed with work from Jianmo Ni [47], which contains the content of reviews and product metadata along with a 5-point rating system. We have used a 5-core subset of the dataset, which means that each of the users and items present in the subset have at least 5 reviews. This ensures a higher quality of reviews. We have performed our analysis on reviews from the Electronics section, extracting only 120,000 reviews which is a reasonable amount.

For purposes of balancing the dataset, we use 40,000 reviews from each category – 'positive', 'negative' and 'neutral'. This is what we use for the 3-class part of the experimentation. We do this by classing 1-2 as negative, 3 is considered neutral and 4-5 is considered as positive. We split this into two parts - 90,000 reviews for training and 30,000 reviews for testing.

For the 2-class variant, we have only two classes- positive and negative. We choose 100,000 reviews from each class (200,000 reviews total) and split this into 2 parts - 150,000 for training and 50,000 for testing.

4.2 Proposed System

In this study, we have listed works that focus on sentiment analysis. We further wish to compare models prepared by adding and removing layers and mechanisms that have been shown in the listed papers to improve accuracy of the classifiers that perform sentiment analysis. Thus, we try to add a CNN layer and an Attention layers. We also demonstrate the use of capsule mechanism to classify reviews. We perform this on LSTM, GRU and Bi-LSTM and Bi-GRU on an Amazon review dataset.

For the 3-class model, we selected 40,000 reviews from the 'negative' class (1-2 stars), 'neutral' class (3 stars) and the 'positive' class (4-5 stars). The total number of reviews were thus 120,000. Regarding the 2-class model, we selected 100,000 reviews from both the 'negative' and 'positive' class, the total being 200,000. Now we performed preprocessing on these reviews, with steps like tokenization, removing punctuation. We retain and use the 20,000 most common, thus not bringing the rare words into account for sentiment analysis. Then the reviews are mapped to a sequence of integers, limiting the length of the sequence to 150. This is because by limiting review length to 150, we get the full text of 90% of the reviews. The rest of the sequence is padded. For preparing the labels of the data, we just turn them into one-hot encodings of dimension 3 for the 3-class model and a single binary variable for the 2-class model.

The models were implemented in Keras, with preprocessing involving word vector embedding using the GloVe pre-processed weights. To prepare the word embeddings matrix, we import the predefined weights from GloVe embeddings data file, and we replicate the vectors of words we have used to the embedding matrix. We use the embeddings of dimension 300, implying that each word is represented by a vector of size 300. This is passed as one of the parameters to the model we create.

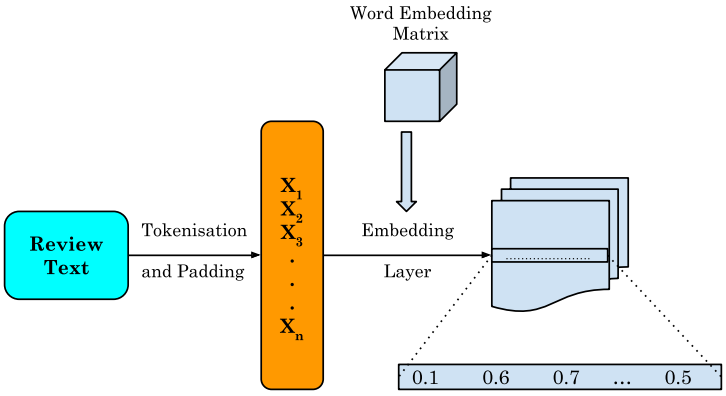

Figure 4.1 Preprocessing steps for reviews

The first layer of all the models is the Embedding layer which contains the embedding matrix we created previously as one of the parameters.

The various models are explained in figure 6. For the models involving the attention and CNN layers, we had the following layers: the Conv1D is the first layer with 128 filters, a kernel size of 5 and 'relu' activation, preceding the unidirectional or bidirectional LSTM / GRU layer, with a dropout regularization of 0.1 applied on the layer. We experimented with different output sizes for the layer, finally deciding on 128. Next, we have a SeqSelfAttention layer that we use with 'sigmoid' activation, and finally we have a Flatten layer to flatten the output. Finally, we have a Dense layer with 3 outputs enhanced by the softmax activation function and categorical cross entropy loss function.

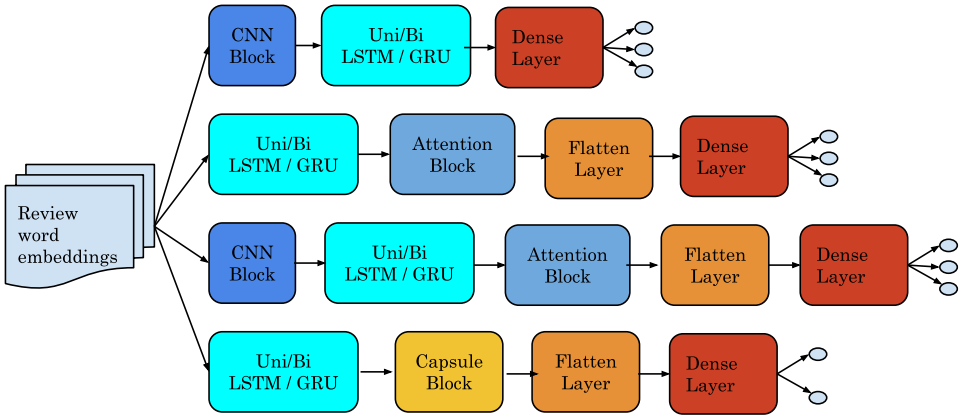

Figure 4.2 Layers for the previous models

For the capsule model, the unidirectional or bidirectional layer is followed by the capsule layer block, with 10 capsules, the dimensions being 16 and the number of routings being 5. Then we flatten the output of this layer, finally having a Dense layer with a single output with 'sigmoid' activation.

We also used early stopping to stop the training after the model performance stops improving on the testing dataset.

4.3 Evaluation Measures

These types of evaluation metrics play an important role in measuring the performance of classification. We have various sampling criteria on the basis of which we can analyze these metrics:

1. TP (true positives) represents numbers of data correctly classified or predicted.
2. FP (false positives) represents numbers of data correctly misclassified or predicted.
3. FN (true negatives) represents numbers of incorrect data classified as correct.
4. TN (true negatives) represents numbers of incorrect data classified or predicted.

Some of the measures that are mostly-used to compare and evaluate the classification method include:

1. Accuracy
2. Precision
3. Recall
4. F-measure.

Accuracy

It is the most common type of metric used for comparing and evaluating the performance of classification. It represents the proportion of the correct predictions of the model. Though accuracy can be a good measure of the effectiveness of a classifier in most cases, it is not enough to give a proper decision. Therefore, precision, recall, and F-measure are to be introduced.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

Precision

Precision is defined as the ratio of all the data that is correctly classified compared to the data that is incorrectly classified. For the variables and features that are in sparse manner, they can be easily classified based on precision.

$$Precision = \frac{TP}{TP+FP} \tag{14}$$

Recall

Recall is defined as the ratio of the actual number of correctly classified samples compared to the one that are correctly classified along with those who were not in a particular category but still, they were not classified. It shows the actual cases of those samples which are not in the required category of the outcome.

$$Recall = \frac{TP}{TP+FN} \tag{15}$$

F1-Score

F1-Score is defined as the perfect criteria for classification as it is a weighted harmonic mean combination of recall and precision.

$$F - measure = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{16}$$

CHAPTER 5 : RESULTS AND ANALYSIS

5.1 Effects of CNN and attention layers

We have tested the effects of convolutional, attention and combinations of them on unidirectional and bidirectional LSTM and GRU networks. There are 4 models for each type of network. The first one is the base model. The second contains the base model in conjunction with CNN layer. The third model contains the base model in combination with self-attention layer. The fourth variant of the network uses both convolutional and self-attention layers in combination with the base model.

In table 5.1, we have compared the different models for LSTM, i.e., the base model, and the base model augmented with convolutional neural networks, attention layers and both of them in tandem. The attention augmented model works best for the 3-class model whereas the base model works best for 2-class model.

Table 5.1 Metrics for LSTM

	3 Class				**2 Class**			
	Acc.	**Prec.**	**Rec.**	**F-sc.**	**Acc.**	**Prec.**	**Rec.**	**F-sc.**
Base	70.05	70.30	70.30	70.30	**93.74**	**94.00**	**94.00**	**94.00**
Base+CNN	73.58	73.33	74.00	73.66	93.47	92.50	92.50	92.00
Base+Att.	**74.91**	**75.00**	**75.00**	**75.00**	92.57	93.50	93.50	93.50
Base+C+A	72.91	73.00	73.00	73.33	92.78	92.50	92.50	93.00

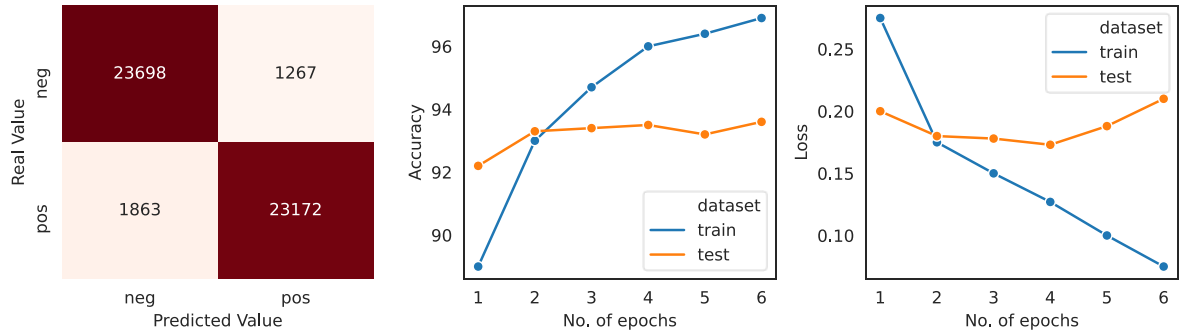

Figure 5.1 Confusion matrix, accuracy, loss graphs for 2 class LSTM

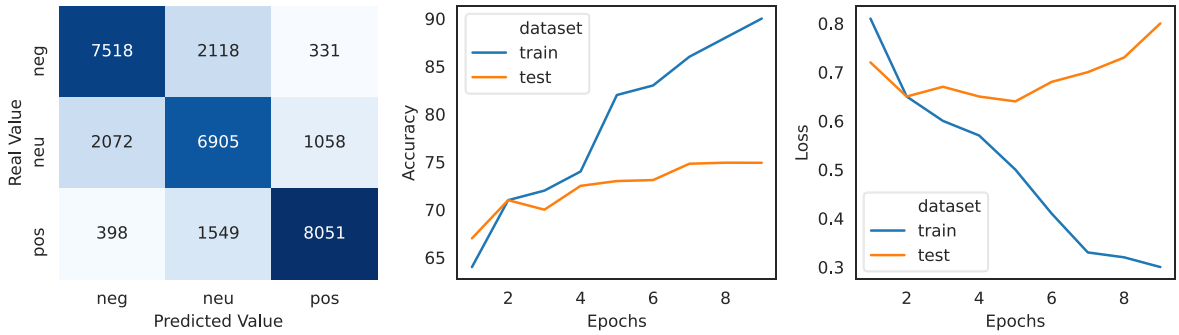

Figure 5.2 Confusion matrix, accuracy, loss graphs for 3 class LSTM with attention

In table 5.2, we have compared the different models for GRU, i.e., the base GRU model, and the GRU model augmented with convolutional neural networks, attention layers and both of them in tandem. The CNN augmented model works best for the 3-class model whereas the base model works best for 2-class model.

Table 5.2. Metrics for GRU

	3 Class				**2 Class**			
	Acc.	**Prec.**	**Rec.**	**F-sc.**	**Acc.**	**Prec.**	**Rec.**	**F-sc.**
Base	71.06	71.30	71.00	71.30	**92.57**	**93.00**	**93.00**	**93.00**
Base+CNN	**72.42**	**73.33**	**73.00**	**73.00**	92.17	92.50	92.00	92.00
Base+Att.	71.23	71.33	71.33	70.33	92.48	92.50	92.50	92.50
Base+C+A	72.09	73.00	73.33	73.00	92.51	92.50	92.50	92.50

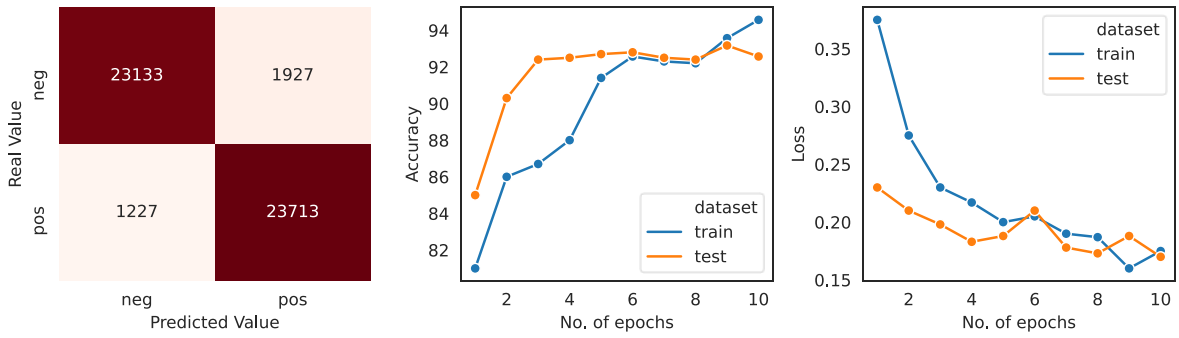

Figure 5.3 Confusion matrix, accuracy, loss graphs for 2 class GRU

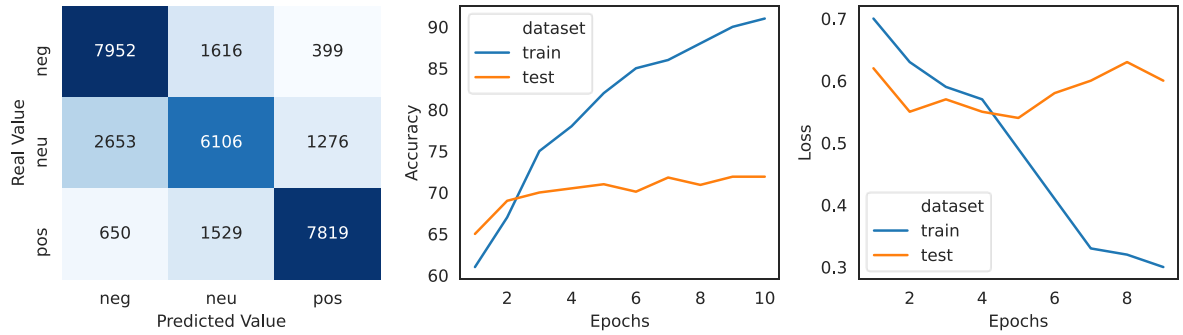

Figure 5.4 Confusion matrix, accuracy, loss graphs for 3 class GRU with CNN

In table 5.3, we have compared the different models for Bi-LSTM, i.e., the base model, and the Bi-LSTM model augmented with convolutional neural networks, attention layers and both of them working together. The attention augmented model works best for the 3-class model whereas the base model works best for 2-class model.

Table 5.3. Metrics for Bi-LSTM

	3 Class				**2 Class**			
	Acc.	**Prec.**	**Rec.**	**F-sc.**	**Acc.**	**Prec.**	**Rec.**	**F-sc.**
Base	70.42	70.30	70.30	70.3	**93.68**	**93.50**	**93.50**	**94.00**
Base+CNN	73.46	73.66	73.33	73.33	92.18	92.50	92.00	92.00
Base+Att.	**74.96**	**74.33**	**75.00**	**75.00**	93.56	94.00	93.50	93.50
Base+C+A	71.75	72.66	72.66	73.00	92.56	92.50	93.00	92.50

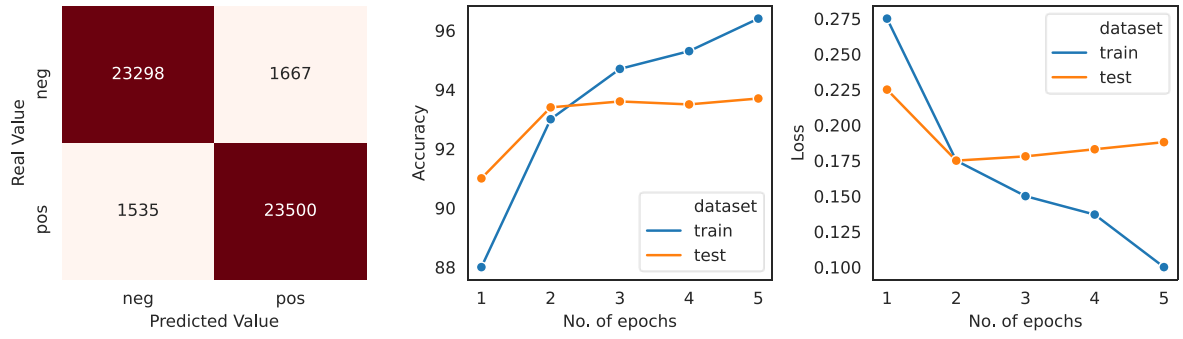

Figure 5.5 Confusion matrix, accuracy, loss graphs for 2 class Bi-LSTM

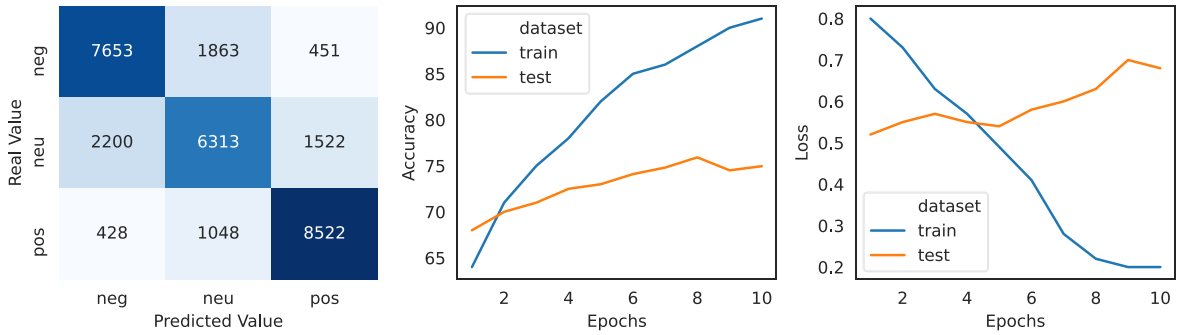

Figure 5.6 Confusion matrix, accuracy, loss graphs for 3 class Bi-LSTM with attention

In table 5.4, we have compared the different models for Bi-GRU, i.e., the base model, and the Bi-GRU model augmented with convolutional neural networks, attention layers and both of them working together. The CNN augmented model works best for the 3-class model whereas the base model works best for 2-class model.

Table 5.4. Metrics for Bi-GRU

	3 Class				**2 Class**			
	Acc.	**Prec.**	**Recall**	**F-sc.**	**Acc.**	**Prec.**	**Recall**	**F-sc.**
Base	71.19	71.00	70.06	70.06	**93.25**	**93.00**	**93.00**	**93.00**
Base+CNN	**72.86**	**73.00**	**73.00**	**73.00**	93.00	93.00	93.00	93.00
Base+Att.	71.00	70.00	70.00	70.00	91.62	92.00	91.50	91.50
Base+C+A	71.00	73.33	70.3	70.2	92.08	92.00	92.00	92.00

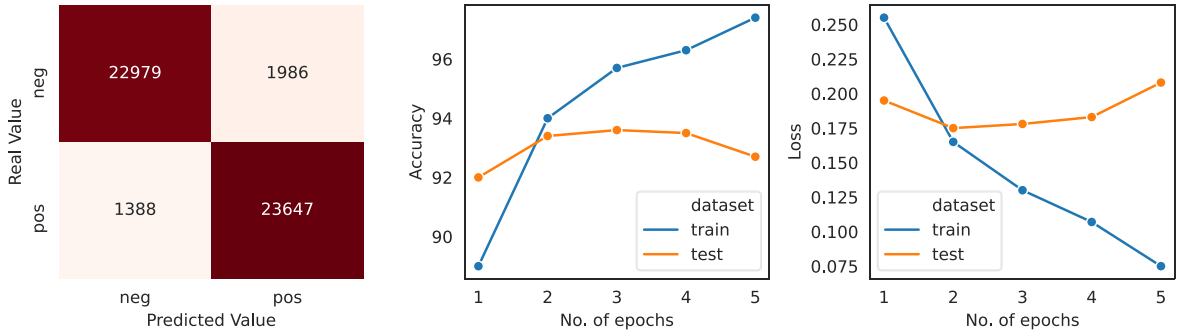

Figure 5.7 Confusion matrix, accuracy, loss graphs for 2 class Bi-GRU

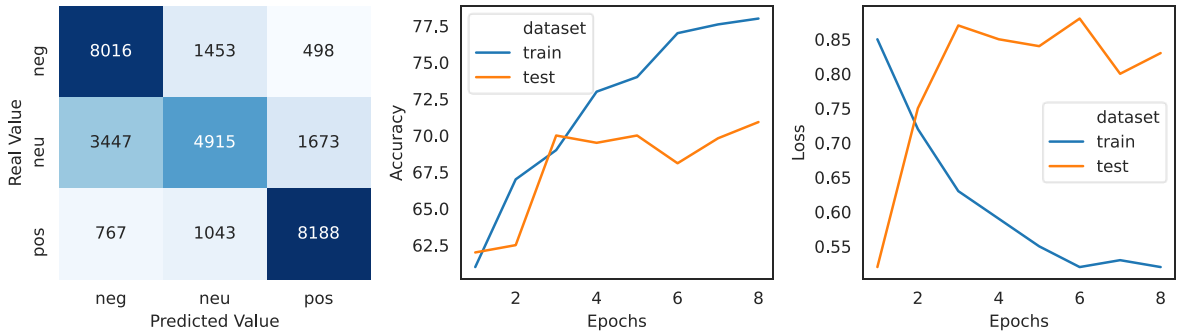

Figure 5.8 Confusion matrix, accuracy, loss graphs for 3 class Bi-GRU with attention

For LSTM, Bi-LSTM and Bi-GRU networks, they appear to work best when augmented by an attention layer. The attention layer is one component of a networks architecture that is in charge of managing and quantifying the interdependence between the input layers. This assigns priority to various values that signify sentiment in a review text and help improve the accuracy of these classifiers.

For GRU network, it is best augmented by a CNN layer before it. This layer convolves the input into an intermediate representation, which contains dimensions each of which can be considered an extracted feature. The GRU then takes these as input. The GRU is a simpler structure with lesser parameters to train, in theory this allows it to generalize better on data, and allows it to better work with the output of the CNN layer.

The 2 class models for LSTM, Bi-LSTM, GRU and Bi-GRU do not seem to improve their accuracy with augmentation of convolutional and self-attention layers. This can be assumed to be because the higher number of layers increases the complexity of the model, which may not have the best setting for optimal results, the simplicity of the base models thus having the highest accuracy.

5.2 Effect of Capsule Network

Table 5.5. Metrics for different models with and without capsule block

		Accuracy	Precision	Recall	F1-score
LSTM	**Base**	93.74	94.00	94.00	94.00
	Base + Capsule	94.18	94.50	94.50	94.50
GRU	**Base**	92.57	93.00	93.00	93.00
	Base + Capsule	93.83	93.50	93.50	94.00
Bi-LSTM	**Base**	93.68	93.50	93.50	94.00
	Base + Capsule	**94.87**	**95.00**	**95.00**	**95.00**
Bi-GRU	**Base**	93.25	93.00	93.00	93.00
	Base + Capsule	93.77	94.00	94.00	94.00

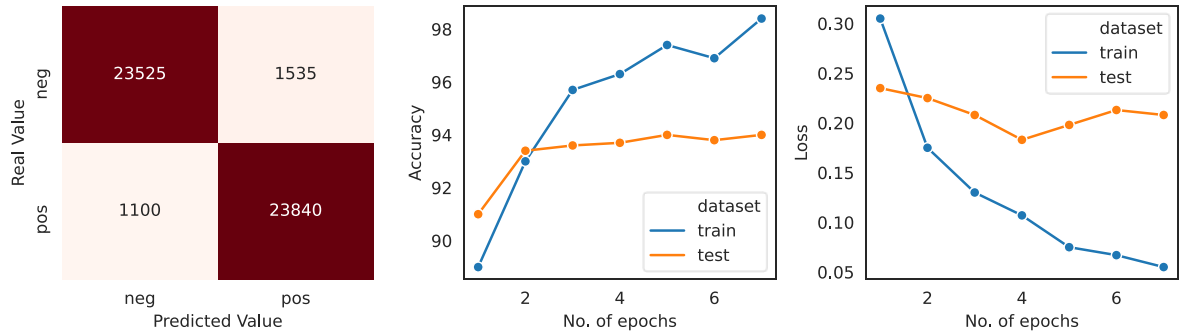

Figure 5.9 Confusion matrix, accuracy, loss graphs for 2 class Bi-LSTM with Capsule

We finally also compare the capsule network by combining it with unidirectional and bidirectional LSTM and GRU, thus finding that it shows the best results with Bi-LSTM. The bidirectional nature of the Bi-LSTM provides a greater potential for learning in the inputs of the capsule block, thus increasing the ability for learning and generalization of the model to obtain a better accuracy in the experiments performed. The Bi-LSTM is the most complex model containing the greatest number of parameters. It thus performs best when used in combination with the capsule block.

CHAPTER 6 : CONCLUSION

After analyzing all these results obtained by us, it is clear that capsule net-work and convolutional and attention layers have their inherent advantages when used in comparison with base models of unidirectional and bidirectional LSTM and GRU. These methods described in this paper can be used to solve a large variety of problems in the fields of both sentiment analysis and deep learning.

Creating understanding from text is a subdiscipline of natural language processing that has many applications. The combination of CNN layers for spatial feature extraction along with LSTM / GRU cell to learn the temporal features is a classic example of the whole being better than the sum of its parts. Attention layers, which help to selectively focus on one aspect of the text while ignoring others can help with this goal as text has inherent structure that makes some parts more important than others. Capsule networks are useful for learning higher level features and details in an object. These combinations of models can be useful for other sentiment tasks. We would desire to see in the future if such models can gather tricky implicit knowledge of human interactions such as humor or sarcasm. We may also aim to assess the performance of these models on data from other domains.

REFERENCES

1. Cambria, E. (2016). Affective computing and sentiment analysis. IEEE Intelligent Systems, 31(2), 102-107.
2. Chaturvedi, S., Mishra, V., & Mishra, N. (2017, September). Sentiment analysis using machine learning for business intelligence. In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) (pp. 2162-2166). IEEE.
3. Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43-54.
4. Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. In International conference on machine learning (pp. 2342-2350).
5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
6. Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)
7. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In Advances in neural information processing systems (pp. 3856-3866).
8. Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: a review. Int J Adv Comput Sci Appl, 8(6), 424.
9. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.
10. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-6). IEEE.
11. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 1480-1489).
12. Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. CoRR Journal, arXiv:1508.01991.
13. Shujie Liu, Nan Yang, Mu Li, and Ming Zhou. 2014. A recursive recurrent neural network for statistical machine translation. In ACL (1), pages 1491–1500.
14. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2008). A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence,31(5), 855-868.
15. Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649). IEEE.
16. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In International conference on learning representations, 2015.
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
18. Hinton, G. E., Krizhevsky, A., & Wang, S. D. (2011, June). Transforming auto-encoders. In International conference on artificial neural networks (pp. 44-51). Springer, Berlin, Heidelberg.
19. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In NIPS 2014 Workshop on Deep Learning, December 2014.
20. Phung VH, Rhee EJ. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences. 2019; 9(21):4500.
21. Chen, Z., & Qian, T. (2019, July). Transfer capsule network for aspect level sentiment classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 547-556).
22. Liang, Y., Meng, F., Zhang, J., Xu, J., Chen, Y., & Zhou, J. (2019). A Novel Aspect-Guided Deep Transition Model for Aspect Based Sentiment Analysis. EMNLP/IJCNLP, 2019.
23. Yang, T., Yin, Q., Yang, L., & Wu, O. (2019). Aspect-based Sentiment Analysis with New Target Representation and Dependency Attention. IEEE Transactions on Affective Computing.
24. Wang, Y., Chen, Q., Ahmed, M., Li, Z., Pan, W., & Liu, H. (2019). Joint Inference for Aspect-level Sentiment Analysis by Deep Neural Networks and Linguistic Hints. IEEE Transactions on Knowledge and Data Engineering.

25. Elnagar, A., Einea, O., & Al-Debsi, R. (2019). Automatic text tagging of Arabic news articles using ensemble deep learning models. In Proceedings of the 3rd International Conference on Natural Language and Speech Processing (pp. 59-66).
26. Haque, T. U., Saber, N. N., & Shah, F. M. (2018, May). Sentiment analysis on large scale Amazon product reviews. In 2018 IEEE International Conference on Innovative Research and Development (ICIRD) (pp. 1-6). IEEE.
27. Mukherjee, A., Mukhopadhyay, S., Panigrahi, P. K., & Goswami, S. (2019, October). Utilization of Oversampling for multiclass sentiment analysis on Amazon Review Dataset. In 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST) (pp. 1-6). IEEE.
28. Ong, R. (2019). Offensive Language Analysis using Deep Learning Architecture. CoRR Journal, arXiv:1903.05280.
29. Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S., & Nagrath, P. (2020). Sentiment Analysis Using Gated Recurrent Neural Networks. SN Computer Science, 1(2), 1-13.
30. Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Generation Computer Systems, 106, 92-104.
31. Zhou, L., & Bian, X. (2019, November). Improved text sentiment classification method based on Bi-GRU-Attention. In Journal of Physics: Conference Series (Vol. 1345, No. 3, p. 032097). IOP Publishing.
32. Zhang, J., Liu, F. A., Xu, W., & Yu, H. (2019). Feature Fusion Text Classification Model Combining CNN and Bi-GRU with Multi-Attention Mechanism. Future Internet, 11(11), 237.
33. Liu, Y., Ji, L., Huang, R., Ming, T., Gao, C., & Zhang, J. (2019). An attention-gated convolutional neural network for sentence classification. Intelligent Data Analysis, 23(5), 1091-1107.
34. Li, W., Qi, F., Tang, M., & Yu, Z. (2020). Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. Neurocomputing.
35. Ambartsoumian, A., & Popowich, F. (2018). Self-Attention: A Better Building Block for Sentiment Analysis Neural Network Classifiers. WASSA@EMNLP, 2019.
36. Ji, L., Gong, P., & Yao, Z. (2019, March). A text sentiment analysis model based on self-attention mechanism. In Proceedings of the 3rd International Conference on High Performance Compilation, Computing and Communications (pp. 33-37).
37. Fentaw, H. W., & Kim, T. H. (2019). Design and investigation of capsule networks for sentence classification. Applied Sciences, 9(11), 2200.
38. Zhao, W., Ye, J., Yang, M., Lei, Z., Zhang, S., & Zhao, Z. (2018). Investigating capsule networks with dynamic routing for text classification. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. (pp. 3110-3119)
39. Yin, H., Liu, P., Zhu, Z., Li, W., & Wang, Q. (2019). Capsule Network with Identifying Transferable Knowledge for Cross-Domain Sentiment Classification. IEEE Access, 7, 153171-153182.
40. Xu, C., Feng, H., Yu, G., Yang, M., Wang, X., & Ao, X. (2019). Discovering Protagonist of Sentiment with Aspect Reconstructed Capsule Network. CoRR Journal, arxiv:1912.10785.
41. Du, Y., Zhao, X., He, M., & Guo, W. (2019). A novel capsule-based hybrid neural network for sentiment classification. IEEE Access, 7, 39321-39328.
42. Zhang, K., Jiao, M., Chen, X., Wang, Z., Liu, B., & Liu, L. (2019). SC-BiCapsNet: A Sentiment Classification Model Based on Bi-Channel Capsule Network. IEEE Access, 7, 171801-171813.
43. Kim, J., Jang, S., Park, E., & Choi, S. (2020). Text classification using capsules. Neurocomputing, 376, 214-221.
44. Zhong, X., Liu, J., Li, L., Chen, S., Lu, W., Dong, Y., ... & Zhong, L. (2019). An emotion classification algorithm based on SPT-CapsNet. Neural Computing and Applications, 1-15.
45. Jain, D. K., Jain, R., Upadhyay, Y., Kathuria, A., & Lan, X. (2019). Deep Refinement: capsule network with attention mechanism-based system for text classification. Neural Computing and Applications, 1-18.
46. Yang, M., Zhao, W., Chen, L., Qu, Q., Zhao, Z., & Shen, Y. (2019). Investigating the transferring capability of capsule networks for text classification. Neural Networks, 118, 247-261.
47. Ni, J., Li, J., & McAuley, J. (2019, November). Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 188-197).
48. Letarte, G., Paradis, F., Giguère, P., & Laviolette, F. (2018, November). Importance of self-attention for sentiment analysis. In Proceedings of the 2018 EMNLP Workshop Blackbox NLP: Analyzing and Interpreting Neural Networks for NLP (pp. 267-275).