Sentiment Analysis using Capsules Network

MSc Research Project
Data Analytics

Shubham Siras Katmusare
Student ID: X19195117

School of Computing
National College of Ireland

Supervisor :
Prof. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Shubham Siras Katmusare

Student ID: x19195117

Programme: MSc. Data Analytics **Year:** 2020-21

Module: Research Project

Supervisor: Prof. Christian Horn

Submission Due Date: 16/08/2021

Project Title: **Sentiment Analysis using Capsules Network**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shubham Katmusare

Date: 16/08/2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

Sentiment Analysis using Capsules Network

Shubham Siras Katmusare
X19195117
MSc in Data Analytics

15th August 2021

Abstract

One of the most important industries on the internet is online merchandise. Nowadays, much of the trade for various products is done online. It is self-evident to conclude that year after year, the internet business's consumer base grows. From the previous year, the number of customers has exploded as a result of the pandemic. In order to meet the needs of the local client base, It is critical to look after the product in the most effective way possible in terms of the quality and the customer satisfaction index for the selling products. The level of quality that these firms are dealing with, as well as their comments, for the items that have been acquired by consumers can lead the organizations to deal with this issue.The purpose of this study is to determine the quality of software goods offered on the online platform "Amazon" as well as consumer satisfaction with product transactions. Sentiment analysis with capsule networks was used to establish the product quality categorization. Based on user feedback, the classification model worked effectively in categorizing product quality as "Good" or "Bad." The study's findings may be used to get business insights on customer happiness and product quality for a certain product category. The capsule network implemented for the sentiment analysis acquired 70.47 percent classification accuracy.

Keywords Deep Neural networks, Capsule networks and Sentiment Analysis.

1 Introduction

In today's society, online purchasing posses a critical role in bringing people from all over the world together and allow them to share their feelings and opinions on issue they are surrounded and concerned with. Amazon.com, for example, is a well-known internet store these days. It allows people to openly post and discuss their thoughts on the company's products. Because of this, an enormous amount of data in structured and unstructured form was produced. Sentiment analysis (SA) can be used to experiment and analyze this data in order to uncover important insights that go hand in hand with people's thoughts and feelings about something which is troublesome and can aid in making choices. Sentiment analysis, is text categorization job that falls under the category of NLP. Machine learning approaches employ Natural Language Processing (NLP) to interpret, analyze, and extract in-depth interpretation from human language (Khurana et al.; 2017). Sentiment Analysis is a computational method related to NLP that allows one to learn about a person's feelings or opinions about a product or an item (Zhao et al.; 2018).

Following the same understanding the proposed research innovates in the field of sentiment analysis on the feedback reviews given by the customers who have bought the software products from the amazon. Most of the research carried out in this filed lacks the specific use cases for the various genres of the data. The Sentiment analysis seeks for the specific type of pre-processing techniques based on the type and field of data (Shah et al.; 2021). To overcome this challenges, the work carried out this research project focuses on specific type of data which concerns with the software products and furnished with the aligned processing steps before carry out the sentiment analysis includes Tokenizing the Data,Stemming and Lemmatization and Feature Extraction. The classification model will be trained based on the data collected from the Amazon user reviews repository and outcomes in the binary classification of the user reviews.

The deep neural network will follow the Capsule networks architecture to support this work. The proposed architecture follows the work carried out by the dynamic routing of the capsules proposed by (Sabour et al.; 2017). Previous work in this subject has done well so far with traditional machine learning methods and pre-trained neural networks for specific tasks. Overall, this research will carry out the sentiment analysis on the user reviews for the software products and will builds a binary classification algorithm based on the contextual information by implementation of the capsule networks. This proposed architecture is less computationally expensive compared to the traditional neural networks which is discussed further in the methodology.

1.1 Research Question and Objectives

"Are capsule networks suitable to extract the context from software product reviews for sentiment analysis ?". The following study objectives have been identified to answer the research question :-

- Investigate scope of deep learning techniques for sentiment analysis of customer reviews.
- Comparison of various word embedding for the research dataset.
- Design a customized capsule network model for classification for user reviews of software products.
- Predictive evaluation of customized capsule network model by the accuracy and recall parameters.

1.2 Research Report Structure

The structure of the research proposal is broken down into the following sections.

1.2.1 Literature Review

The past work pertaining to the research statement is included in this section. The six segments take a step-by-step look at the literature. The Amazon review data is briefly described, as well as the related work in the field of sentiment analysis utilizing machine learning and neural networks. The simulation of a capsule network is described after accepting the facts regarding capsules in neural networks. In this part, you'll learn about the capsule network's benefits and drawbacks, as well as the elements that influence its performance.

1.2.2 Research Methodology

The methodological approach to attaining the research idea mentioned in this study is presented in this part. The approach to the research objectives is discussed in detail, including subjects such as research needs, data collecting, various pre-processing techniques to use, and transformation required. Finally, a short on the simulation of capsule networks for sentiment analysis concludes the methodological overview along with the factors affecting its performance.

1.2.3 Design Architecture and Implementation

This section covers the design architecture as well as the needed specifications. The exact architectural design, network structure, neural network outlines proposed, and convolutional layers involved are all briefly discussed so that the goal may be realized by following the architecture flow. A overview of the design strategy is also provided, along with justifications for each item of work proposed. This section discusses the main reasons for selecting each unit in the architecture. Further, the implementations strategies and the exact work is reflected in this section. Finally, section will summarize the actual summary of the artifacts for this project.

1.2.4 Experiment and Evaluation

Section starts with the briefs of data splitting and statistics of data which have been used for training. Secondly, the various parameters with their specific values which have have been in research have been discussed in detail. This section also conclude and compares the outcomes of this research work with the existing work on this dataset. The purpose of section is to give the conclusive reflection of the research work carried out with its advantages and disadvantages.

1.2.5 Conclusion and Future Work

The aim of the objective stated in this article is concluded in this piece of research report. The direct outcomes that may be expected with the suggested design, as well as the future scope in various applications, are described in this section.

2 Literature Review

2.1 Introduction

Many researchers have been researching in this subject in recent years. Many concerns with slangs, non-context characteristics, and temporal conditions have been raised by research. Some researchers have even explored on languages that are multilingual. The brief text data has been studied by researchers. Few researchers have indeed looked at utilizing various keywords to express context of text. It analyzes people's views using textual data accessible nearly anywhere, including Amazon, and focuses on the topic portion of the text that results in a good or negative context (Liu; 2015). Sentiment analysis is critical in the business world because it gives companies an in-depth understanding of how customers feel about their product, allowing them to adjust their strategy to fulfill the customer's expectations and requirements while avoiding loss. On the other side, it is beneficial to potential buyers in deciding which items to purchase (Zhang and Liu; 2017).

With the another perspective, At the document level, there has been a significant amount of effort put on sentiment categorization of reviews on online businesses using machine learning approaches (Pang et al.; 2002) and (Athanasiou and Maragoudakis; 2017). In this approaches, the model breaks down a document containing reviews into small chunks that is sentences, then analyses each phrase for structure and contextual reliance of each entity inside the review text to identify the sentence's emotion orientation (Xu et al.; 2012). The majority of research in sentiment analysis for product user reviews use binary classification, in which reviews are categorized as "good" or "negative." Furthermore, even the greatest systems presently only achieve an F1-score, precision, and accuracy of approximately 80 percent (Saqib et al.; 2018).

At the document level, there hasn't been quite enough work on the same issues utilizing lexicon-based approaches. Nevertheless, there has lately been progress in the development of lexicons for sentiment analysis. Amazon allows customers to evaluate products on a scale of one to five stars (considering 1 low and 5 as maximum), as well as offer a written summary of their experiences and views concerning the brand and the supplier. This ranking system is used to designate the text reviews. In the data, reviews with one, two, or three stars from customers considered as low level products and are classified as 'negative' score, whereas those with four or five stars are considered as best and good quality products and labeled as 'positive'. The text reviews are converted to term Frequency Inverse Document Frequenc (TFIDF) feature vectors made up of all the lexical items in the reviews. TFIDF scores are included in each of these feature vectors. A TFIDF score is summation of its frequency and relative significance inside a document. The proposed architecture uses pre-processed texts after converting them into the word embeddings processed through TFIDF algorithm.

conclusively, to assess the polarity of customers toward software goods, we need an efficient and comprehensive sentiment analysis framework for the provided dataset. Capsule network will be used to create this design. Because the proposed study focuses on sentiment analysis and categorization of customer reviews, we'll go through some of recent and relevant work in this genre of Amazon user review data and sentiment analysis utilizing machine learning and deep learning approaches.

This section is categorized as following sections. The first subsection 2.2 briefs about the NLP and its link with the sentiment analysis. This section explains about various libraries which can be used to implement and carry out the sentiment analysis tasks. Eventually, the section is concluded by discussing the lexical techniques in field of machine learning. The section 2.3 describes about the machine learning work carried out in field of sentiment analysis. we have discussed various vectorization methods and earlier state of art work carried out earlier in this context. The upgrade in sentiment analysis is actually achieved by implementing the neural network. The capability of understanding the contextual information using neural networks have outperformed and such studies are referenced here. furthermore, section 2.4 covers the aspects and scope of capsule network in filed of sentiment analysis as it is well known in filed of image classification. we have also discussed how to simulate the capsules and further implement it using its three basic concepts that is transforming auto-encoder, dynamic routing of the capsules and expectation Maximization(EM) Routing and Matrix Capsules. Finally in the last section

2.6, we have have noted down which are all factor which needs more focus while implementing capsules because of its high influence to the output.

2.2 Natural Language Processing (NLP) and Sentiment Analysis Background

2.2.1 Natural Language Processing

Unstructured text makes up a large portion of user-generated material. Because of the large volume of unstructured data, a set of machine-based ways enabling computers to analyze material and comprehend language has been developed. In this article, we employ a variety of NLP techniques to analyze the review comment fields and convert them into machine-readable vectors. Natural Language Toolkit (NLTK) and SpaCy are among the Python packages we used. NLTK includes more than 50 text and lexical resource collections, as well as numerous tools, interfaces, and techniques for processing and analyzing text data. SpaCy is a new package that implements each NLP method and algorithm to the best of its ability. To convert text data into its vector form and implementing machine learning features available in Python such as Pandas, NumPy and SciPy stack libraries are utilized.

2.2.2 Sentiment Analysis

Sentiment analysis is a text categorization discipline. Sentiment analysis is the practice of utilizing NLP and text analysis technologies to discover and extract subjective information from a piece of text. Sentiment analysis performs better when there is a subjective context to the text than when there is simply an objective context. The reason behind this is, whenever a body of text content has an context or viewpoint for certain objective, it typically portrays some ordinary facts without conveying specific emotion, sentiments. Subjective text is typically expressed by an individual in a particular mood, emotion, or feeling.

2.2.3 Machine Learning Techniques

Sentiment analysis difficulties are tackled using a variety of approaches. Unsupervised machine learning is a type of approach that utilizes classification algorithms to categorize items based on their related sentiment. Learning using existing set of training data points is baseline for supervised learning. Logistic regression (LR) and support vector machine (SVM) are two frequently used supervised machine learning methods for text categorization (Joachims; 1998).

The logistic function, commonly known as the logistic regression method, is a classification technique that assigns data to a discrete set of classes. For classification problems, LR can be a reliable approach. SVM is a supervised learning approach that divides data into two sets implementing hyper-planes. SVM is one of the most implemented classifiers in recent years. Studies have extensively used the Gradient Boosting machine learning approach for text classification and deduced that it outperforms SVM and LR. Gradient Boosting machine learning is a simple decision tree-based method. Each round of Gradient Boosting training fits a newly tuned model to improve classification. The loss function's negative gradient is associated with each new model, and the loss is reduced via gradient descent (Chen and Guestrin; 2016).

2.3 Machine Learning Approach for Sentiment Analysis

The problem of extracting the best and most accurate features and simultaneously categorizing the customers reviews into good and bad thoughts has become a prominent study field. A study, first to proposed sentiment categorization employing machine learning techniques on a movie reviews dataset appeared in below mentioned research paper. For sentiment analysis, they looked at the Support Vector Machine(SVM) and Naive Bayes(NB) models. SVM implemneted using unigram feature extraction apprroach gave the best results in the trial. They achieved an 82.9 accuracy (Pang et al.; 2002).

Mullen studied sentiment classification on apparel, shoe and jewelry items customers review datasets (Mullen and Collier; 2004). They evaluated feature retrieval approaches based on Lemmas with hybrid SVM, Nave Bayes, LR, and decision tree algorithms. SVM generated the greatest results in their analysis, with an accuracy of 86.6 percent. Lilleberg published a study which compares Word2vec and TFIDF text feature extractions by implementing SVM. They also verified the outcomes of categorization considering and avoiding the stopwords. The highest SVM performance they found with TFIDF and excluding stopwords was 88.9 percent accuracy matrix (Lilleberg et al.; 2015).

SVM and LR have become popular classification algorithms for document analysis in recent years. Elmurngi and Gherbi proposed using SVM and sentiment analysis to prevent fraudulent movie reviews. The text containing stopwords and without stopwords, they evaluate the results of SVMs, Nave Bayes, decision trees, and KNN classifications. SVM was the most accurate in both situations, with accuracies of 81.75 percent and 81.35 percent, respectively (Elmurngi and Gherbi; 2017). In a separate paper, Ramadhan performed sentiment analysis on a social media Twitter dataset implementing logistic regression via TFIDF feature extraction (Ramadhan et al.; 2017). It was stated that the categorization accuracy was near to 83 percent. On an Amazon product review dataset, a combined study of various Machine learning models performed an experiment utilizing SVM, TFIDF model, and Next Word Negation which reported accuracy of 88.86 percent (Das and Chakraborty; 2018).

Rodrigues and Bhavitha published a paper that compared different machine learning approaches on movie reviews. For the SentiWordNet technique, they reported an accuracy of 74 percent and for SVM they achieved 86.40 percent (Bhavitha et al.; 2017). Athanasiou used machine learning featured by Gradient Boosting for sentiment analysis and discovered that it outperformed SVM, Naive Bayes, and neural networks for both balanced as well as unbalanced data sets. With an accuracy of 88.20 percent, the Gradient Boosting machine learning fared best (Athanasiou and Maragoudakis; 2017).

2.4 Sentiment Analysis by Neural Networks

Neural network-based models proved and demonstrated tremendous credibility in various NLP problems from the development of a successful technique to acquire and learn distributed representations text data (Mikolov et al.; 2013). Recursive Auto Encoder (Qian et al.; 2015a), Recursive Neural Tensor Network (Socher et al.; 2013), Recurrent Neural Network (Tang et al.; 2015), Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber; 1997), Tree-LSTMs (Tai et al.; 2015a), and Gated Recurrent Unit (GRU) (Cho et al.; 2014) are only a few of the models used in sentiment analysis.

The representation of a sentence is built iteratively using a recursive autoencoder neural network (Qian et al.; 2015b). In most cases, recursive models rely on a tree structure of input text. All sub phrases must be annotated in order to achieve competitive outcomes. Tree-basedLSTMs have shown to be successful for a variety of NLP applications, including sentiment analysis (Tai et al.; 2015b), by using phrase syntactic structures. Tree-structured data is not required for sequence models like CNN, which are frequently used for sentiment classification (Kalchbrenner et al.; 2014). Because of its capacity to model the context of text data, LSTM is indeed commonly used to develop sentence-level representation. Regarding the effectiveness of those approaches, discriminating between various emotion polarity remains difficult.

The suggested neural network architecture enhances learning using this neural word embeddings to leverage the distribution of word co-occurrences . Each inferred aspect's top representing words represents the aspect, resulting in more relevant outcomes. The prediction may be made using only the retrieved short and coherent chunks of text, which can also be utilized to explain the forecast.

2.5 Simulation of Capsule Networks

Capsule networks are a form of CNN with one distinguishing feature known as "Capsules". Capsules are a collection of neurons. The capsule represents a vector input with orientation characteristics, whereas the neuron's output is a scalar number. This gives you an advantage in capturing the entire target using its sub-parts. The kind of capsules assigned determines how the object or entity's features are handled. The capsule's output contains the likelihood of the feature's existence as well as a collection of vector values known as instantiation parameters. The network's capacity to verify posture, deformations, and texture is shown by these parameters. The capsule networks are characterized by three separate methods: dynamic routing based on capsules, transforming auto-encoders, and matrix capsules, all of which have been addressed in the literature further below.

2.5.1 Transforming Auto-encoders

The transforming auto encoders were the starting point for capsule network research. The main goal of this search was not to focus on object identification, but rather to get a test orientation of features and its characteristics from which to generate a output. The idea behind this method is to organize the capsule in different layers (Hinton et al.; 2011). Consider one level of capsules at the primary level and a second level of capsules at the secondary level. The primary level capsules will detect and gather the entity's posture and orientation characteristics, as well as begin the hierarchy's flow. Lower level capsules should be interconnected with higher level capsules using inter specific connections in order to follow the hierarchy. This entire hierarchy is advantageous for extracting orientation parameters for entity pieces and passing them on to higher secondary level capsules to obtain the entire entity. The feature-by-feature learning will eventually be completed if this procedure is followed.

2.5.2 Dynamic Routing of the Capsules

The group of neurons with the instantiation parameters is the next step of change in capsule networks. The nicest part about this method is that no positional or orientation parameters are required as input. Convolutional layer, main capsules, and Class capsule layer are all part of the suggested method. The convolutional layer extracts entity features, which are then sent to the class capsule layer. The capsules will record a variety of essential data related to the item, to get the complete context of the input. The length of the activity vector would be used to determine the scope of the object's existence, and the orientation parameters are used to determine the instantiation parameters (Sabour et al.; 2017). Lower-level capsules will approaching higher-level capsules, resulting in greater activity vector scalar product magnitudes. This method allowed for a better understanding of the output or short context of features.

2.5.3 Expectation Maximization(EM) Routing and Matrix Capsules

The combination of matrix capsules with EM routing instead of utilizing vector outputs, (Hinton et al.; 2018) provided a new approach to investigate the depths of capsule networks. It was necessary to lower the dimension of the transformation matrices in between capsules from n*2 to n to lower calculation cost. So, adoption of the EM routing algorithm allowed the capsule network to function well at several layers. This project focuses on feature space and orientation from various angles or perspectives. A capsule is a cluster of numerous neurons with different characteristics.Every capsule throughout the network layers does indeed have a logistic unit with a 4*4 matrix that contains the locations of entities from the perspective of the observer. The implementation of a novel iterating routing technique based on the EM algorithm allows lower-layer capsules to cluster with higher-level capsules that have comparable orientation information.

2.6 Factors Affecting Performance of Capsule Network

The performance of an algorithm is determined by the qualities of any dataset. When opposed to numerous complicated datasets with noise and inconsistencies, a simple dataset may just have one stream to focus on. Despite the complexity of the datasets, numerous studies have shown that capsule networks perform better than standard machine learning methods and CNN. Momentum, batch size, learning rate decay, and drop out rate have little effect on the performance of capsule networks, while routing operations and the frequency of routing operations have a considerable impact (Lin et al.; 2019). After optimizing the amount of routing operations, (Ren et al.; 2019) found an enhancement in the robustness of capsule networks and inferred that networks converge quicker under Adam and Rmsprop optimization.

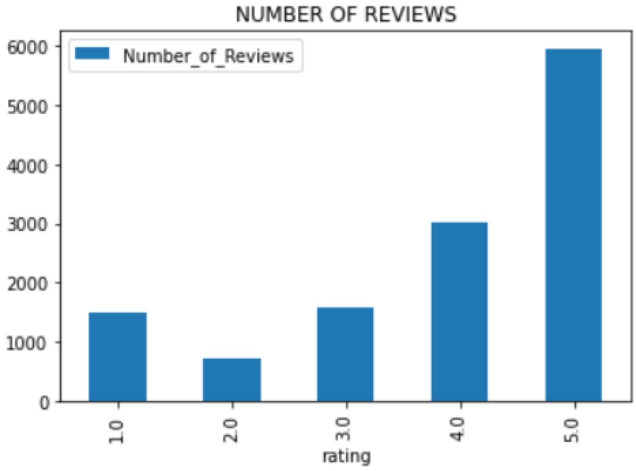
3 Research Methodology

This section explains the benefits and steps involved in establishing a novel method based on capsule networks. In this study article, the suggested approach for sentiment analysis for amazon user reviews which will result in binary classification utilizing capsule neural networks is based on the Knowledge Discovery in Database (KDD) technique.

3.1 Data Selection and Understanding

Various items Consumers want and then provide feedback in the form of a comment or a graded technique ranging from 1 to 5 stars, among other options. The dataset was retrieved from the Amazon user reviews public repository with the agreement that it will only be used for study and not for commercial purposes. Product ID, overall rating, customer varified, reviewTime, reviewerID, reviewerName, review text and product description are all included in the dataset. The data is in JSON format. most of the fields were omitted in order to make the dataset ready for the study which were not text type. Following are the distribution of the ratings from the various users from range 1 to 5. This will give the brief about the classes we are going to use for classification using sentiment analysis. The figure below justifies the distribution of the various user reviews ratings over the considered dataset. This user ratings are eventually be reduced to two classes which in discussed due further in implementation section.

Figure 1: Distribution of Reviews



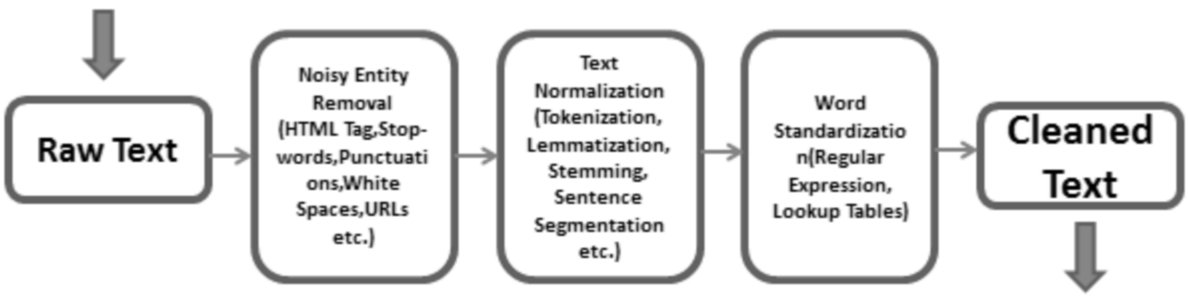
3.2 Data Preprocessing and Transformation

The importance of text preparation in sentiment analysis cannot be overstated. Because a trained ML model will perform well if its is trained on good and featured input data, considering text preparation is given specific attention (Athanasiou and Maragoudakis; 2017). The majority of user-generated material is unstructured. As a result, in order to train any ML model with certain data, some procedures must be taken to normalize the input. The following are the key steps involved in text normalization for proposed research:

3.2.1 Perform Review Extraction

In this step, we read the JSON file which was available on the amazon user reviews repository. The data in JSON format possessed several data-points holding various fields namely Product ID, overall rating, customer varified, reviewTime, reviewerID, reviewerName, review text. after performing some file handling techniques we finally retrieved the columns which are significant for this research problem. Finally, reviewer id, user rating and review text were considered for the further analysis (Aliases - 'reviewerID', 'rating', 'reviewText' were used to represent this fields). Some of the processing and transformation carried out contextualised in the flowchart below which have been implemented in this research.

Figure 2: Data Flowchart for Preprocessing and Transformation



3.2.2 Remove Special and Accented Characters

Accent touch in the text data cause a lot of problems in data cleaning and preprocessing techniques. For an instance, it affects lemmatizing and expanding contractions, therefore we can delete them as a precaution. In this stage, we also delete any additional special characters or emojis.

3.2.3 Stopwords Removal

Stop words are words like "I," "to," and "the" that have little meaning. This is used to finish the phrase and has no influence on the contextual meaning of the statement. For eliminating such terms from the corpus, we employ the NLTK stopwords, punkt, and wordnet corpus.

3.2.4 Text Lemmatization

The Strategy or method of finding a linguistic root of words is known as lemmatization. Lemmatizing allows algorithms to detect various tenses from same text in many situations. The WordNet lemmatizer package from the NLTK library is used in our preprocessing steps of the research.

3.3 Vectorization

After performing the preprocessing steps mentioned above, the data is ready to get vectorized. Feature extraction is the process of extracting useful qualities from raw textual data before feeding it into a statistical or machine learning method. The ultimate result from this technique is sequential numerical vectors, hence also known as vectorization. Because traditional algorithms operate with numerical vectors, they can't work directly with raw text data. We chose to utilize the TFIDF approach in this research since it is widely regarded as the one of the best feature extraction technique in text analytics.

The frequency of recurrence of set of terms in a document is referred to as TF. This approach is also known as the Bag of Words model. Each document is parameterized of 0s and 1s in this paradigm. If a word appears in a document, its associated vector location will be recorded as a "1" and if word is absent, it recorded as a "0".The following is how TF is formulated:

Figure 3: Term Frequency (TF)

$$\text{TF(word)} = \frac{\text{Frequency of Word in the Document}}{\text{Number of Word in the Document}}.$$

The IDF for a word can be define as a representation of how important that term is across the entire corpus. The following is how the IDF is formulated:

Figure 4: Inverse Document Frequency(IDF)

$$\text{IDF(word)} = \log\left(\frac{\text{Total Number of Documents}}{\text{Number of Documents Containing the Word}}\right)$$

For specific words, TFIDF is the multiplication result of TF and IDF scores. Each document in the TFIDF model resembles as a vector containing TFIDF scores for each and every words in a text. The TFIDF reduces the affect of frequent but ineffective characteristics of words. In this research, we use the Python Scikit-learn library's TfidfVectorizer and TfidfTransformer modules to create our TFIDF model. On the text data, these fit and transform features. Only unigrams are included in the vectorized TFIDF (single words).

3.4 Data Mining using Capsule Networks

Deep neural networks have seen a lot of success in field of sentiment analysis in past few years. Nevertheless, unlike humans, these deep models were data-hungry and generalize poorly with limited samples. Capsule networks have improved the generalization ability of deep networks in image classification. They may generalize to the same item in multiple 3D pictures from different perspectives and view points. Extrapolation can be used to develop such generalizability from instances with few views points (Hinton et al.; 2011). This implies that capsule networks in NLP applications should abstracts away from various surface realizations in a similar way. Capsule networks have the capacity to learn hierarchical connections between successive layers utilizing no-parameter routing procedures, which are clustering-like techniques that increase generalization capabilities. Such routing techniques are contrasted with pooling and fully connected layers.

The development of capsule networks for mature NLP applications is currently hampered by a few main obstacles. For instance, routing procedures, rely on iterative routing low-level capsules to high-level capsules in order to better understand hierarchical links across layers, thus choosing the number of iterations is critical. Existing routing methods, on the other hand, employ the same amount of iterations for all cases, making it difficult to assess routing convergence. At the system level, a routing method with five iterations on all examples converges to a reduced training loss, but not at the instance level for one example. Due to the huge number of capsules and possibly vast output spaces, training capsule networks is somewhat more challenging than standard neural networks like CNN and LSTM, and necessitates substantial computing resources in the routing. In this research we will try to overcome this challenges by implementing own softmax function.

3.5 Model Evaluations

True positives, false positives, true negatives, and false negatives are the outcomes of binary classification for proposed classification model in this research. False negatives and false positives are considered as predictions which are mis-classified, whereas true positives and true negatives properly anticipate real datapoint labels. These terms are discussed further in brief.

- True Positive (TP) : This is a sample of the Number of positive instances that were correctly anticipated.
- False Positive (FP): This is a sample of the Number of positive instances that were incorrectly anticipated.
- False Negative (FN): This is a number of negative instances incorrectly categorized.
- True Negative (TN) : This is a number of properly identified percentage negative instances.

Precisely, we are targeting to achieve the model evaluations based on the accuracy and recall matrices which are solely based on the outcome of confusion matrix. These matrices are briefly discussed further.

3.5.1 Accuracy

The classifier's accuracy measures how well it makes the correct prediction. The fraction of the precise instances of predictions to the total instances of forecasts is known as accuracy. It is expressed as below.

Figure 5: Accuracy

$$\textit{\textbf{Accuracy}} = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}}$$

3.5.2 Recall (R)

It determines the sensitivity of a classifier, or how much optimism it delivers. High remembrance indicates fewer false negatives. The recall is the proportion of correct positive classified instances to the total number of expected positive class instances. It is formulated as below.

Figure 6: Recall

$$\textit{\textbf{Recall}} = \frac{\text{TP}}{\text{TP + FN}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{Actual Results}}$$

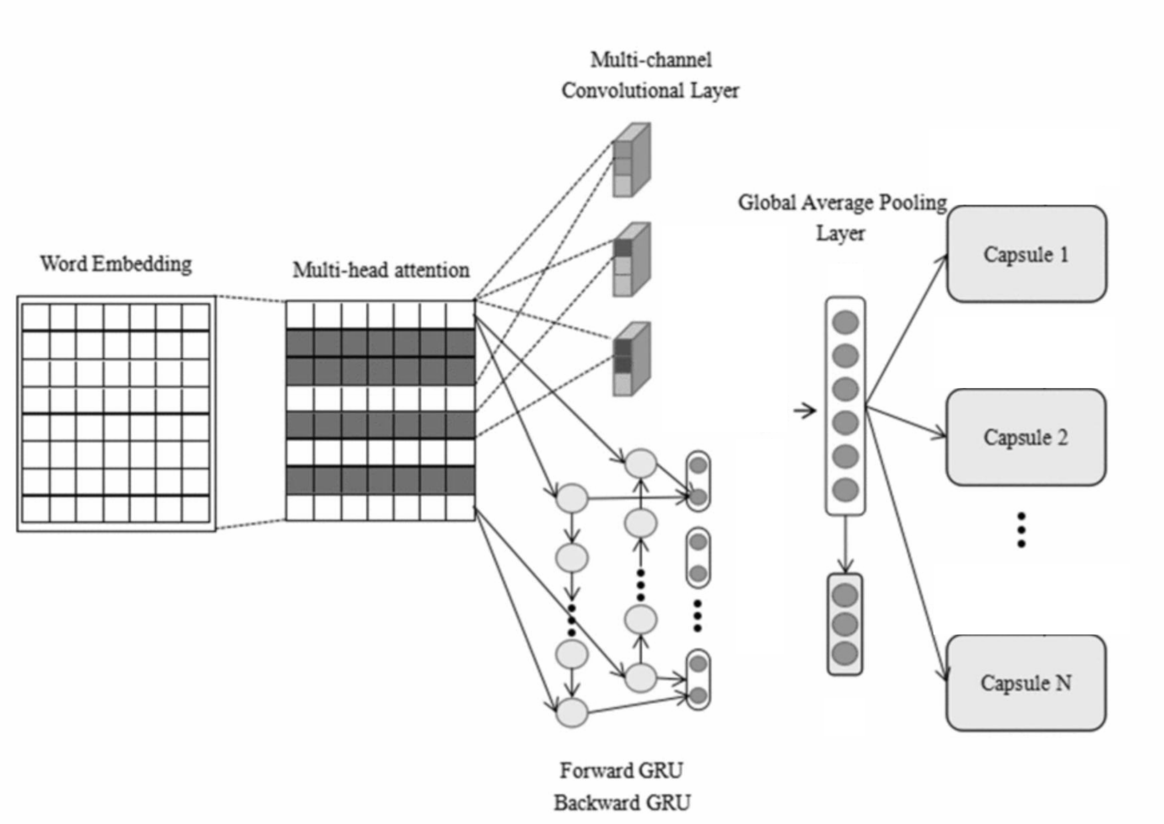
4 Design Architecture and Implementation

This section outlines the design principles that will be used to achieve the goal, as well as the network architecture that will be used throughout the implementation phase.

4.1 Proposed Architecture

The Capsule network's model framework for the proposed research is depicted in Figure below. Text encode layer, text converge layer, text extraction layer and capsule construction layer are the four elements of the model suggested in this article.

Figure 7: Capsule Network using Bidirectional Gated Recurrent Unit for Sentiment Analysis



The Text encode layer consists of a multi-head attention mechanism that collects features from the text, captures word relationships, and creates a text feature vectors.Input feature word vectors into CNN and Bi-GRU consecutively, depending on multi-head attention output. After that, the convolution operation is splicing is done. The goal focus on to get the N-gram characteristics of a every text's words and pass them into the model's subsequent layer. As a result, only the convolutional layer will be utilized to get local text features, whereas a Bi-GRU model processes text and extracts global features using the forward directional GRU and backward GRU. Using the text converge pooling layer to get feature so as to obtain the features annotations of the review text in order to compute the loss function, dividing the retrieved global and local features to produce the feature matrix to feed as inputs of the featured capsule. The target categorization category corresponds to the number of feature capsules. Two capsules, for example, correlate to pleasant sentiment and a bad evaluation. Capsule layer, A capsule is deemed activated if its activation probability is the highest among all capsules; else, it is passive. The target categorization of the input data as the response from model is the attribute pertaining to the activated characteristic of capsule. So implementing the capsule network using proposed units, the objectives are tried to achieve.

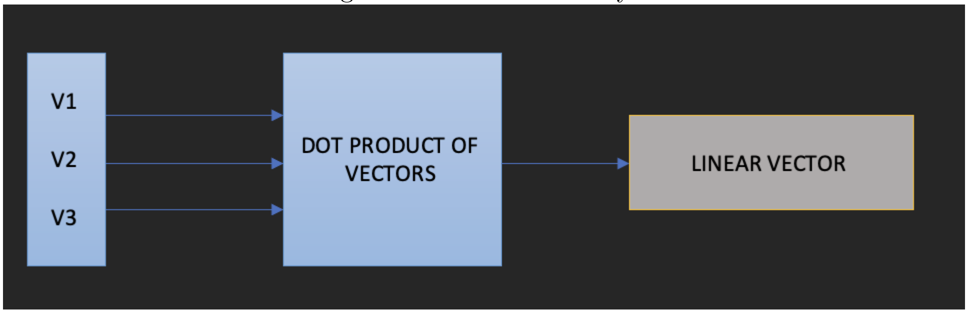
4.2 Network Structure

The proposed network in this research seeks for the understanding of the 4 units which altogether serves the objective of binary classification for proposed sentiment analysis. The proposed network can be elaborated in the four different parts listed below.

4.2.1 Text Encode Layer

The attention mechanism has the ability to selectively focus on the text's most significant information. As demonstrated in Figure below, the proposed research employs multi-head attention in order to collect significant features of the text data from several sub spaces.

Figure 8: Text Encode Layer

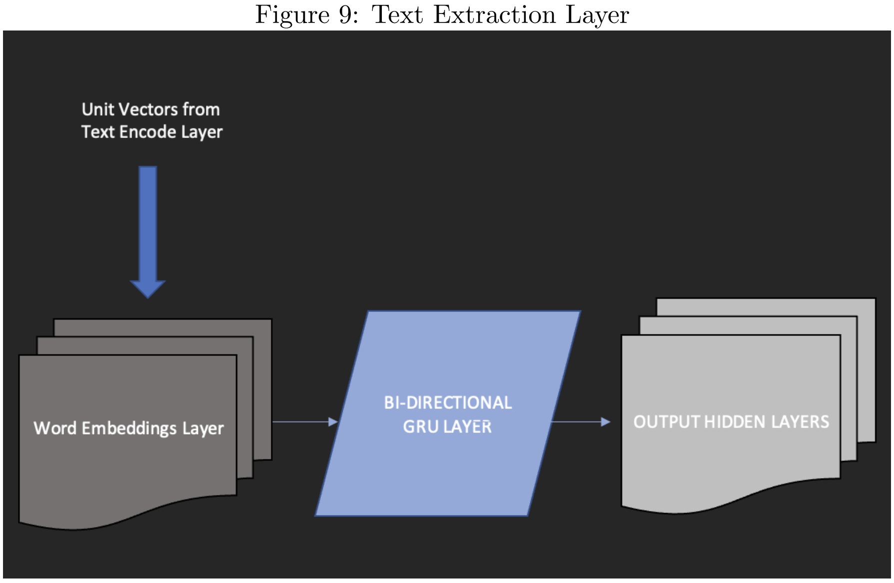


Every entity in a input review will be mapped to a vector of N-dimensions. The word vector matrix will be first linearly converted and split into three matrices of the same dimensions, each of which is assigned to a separate subspace. Then, for each of the three matrices, compute the value of attention of every subspace in simultaneously. Sub space's attention value converts the attention matrix to regular normal distribution in order to avoid gradient loss which is being lost in the reverse propagation procedure. After that, each sub space's attention value is sliced and linearly converted.

4.2.2 Text Extraction Layer

This research utilised the benefits of Bi-GRU for text features extraction to model the text context characteristics in order to retrieve more complete features. The Recurrent Neural Network (RNN) possesses the capacity to integrate every lexicons in the corpus into to the model's learning, not like traditional approaches which consider only specific corpus context as the implicit features of sentiment analysis algorithm.

Figure 9: Text Extraction Layer



Standard RNN, on the other hand, have the issue of disappearing or bursting gradients. To address the issues, GRU and LSTM follows construction of gates to utilise features in order to specifically alter various characteristic of every instant of model. GRU is a version of LSTM that uses update gates instead of the forget and input gates. The state transmission in a traditional recurrent neural network is unidirectional. In some circumstances, however, the recent output is not just connected to the prior state, and sometimes even not to the consecutive next state. For an instance, anticipating the absent words in a text necessitates not only the prior judgment but also the information of the later lexicons, an issue that was solved by the development of the bidirectional recurrent neural network. Two unidirectional recurrent neural networks are combined in the bidirectional recurrent neural network. At any one time, input is sent in opposing directions to two RNNs, and the outcome is jointly decided by them, resulting in a more accurate result. To learn global semantics result matrix, model proposed in the research employs a Bi-GRU network. During the training phase, the network utilizes two GRUs at the very same instant to simulate features along the front and backward of the embeddings, and then generates the hidden layers.

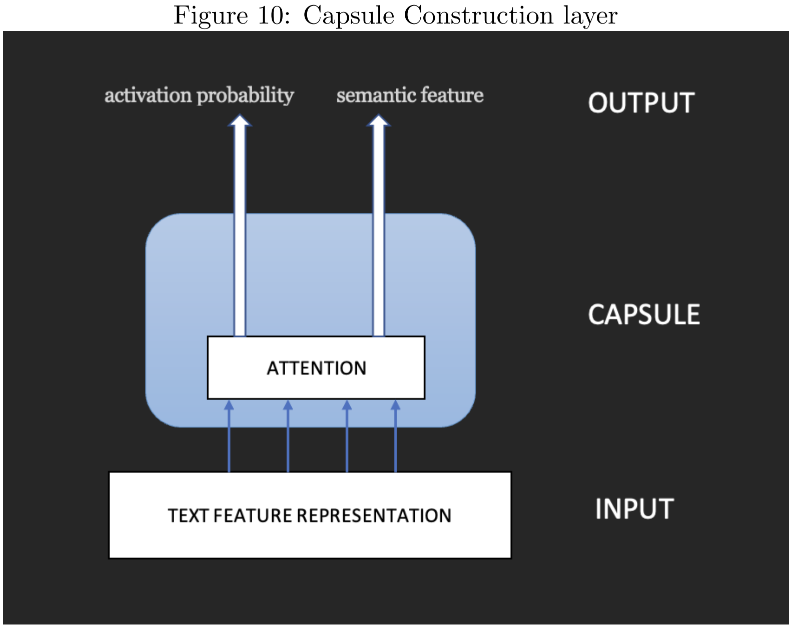
4.2.3 Feature Converge Layer

When retrieving local text features, the CNN decreases information loss. To retrieve global semantic characteristics, the Bi-GRU network gets through the whole input review text. The average pooling approach is used to combine the global and local features of the review text data which helps in creating the feature vector, this improvise the sentiment extraction and feature expression capacity. while carrying out the research, convolution kernels quantity in CNN and the dimensions of bidirectional GRU network's output vector are considered on same magnitude, and the feature map produced by the these networks is spliced by implementing splicing and merge technique. The pooling layer averaged the feature vectors to produce feature points. The representation of features of such sentiment instance is made up of these concluded feature points into the final feature map, which prevents over-fitting and improves the model's resilience.

4.2.4 Capsule Construction Layer

We can retrieve the final feature capsule after traversing from above mentioned layers. The capsule feature representation is constructed using the attention mechanism; the capsule activation likelihood is predicted using activation function called sigmoid; and the reconstructed feature representation of the capsule is obtained by multiplying feature representation and activation probability. The attention mechanism had previously used in machine translation jobs, and the representation module constructs the capsule's feature representation with the attention mechanism by combining the pooled feature vector. The presenting module can use the attention mechanism to determine the significance of words in various texts. For instance, in hotel review data, the word "spacious" gives good information, but its value in movie reviews is diminished.

Figure 10: Capsule Construction layer



The capsule's three components are complementary to one another. Each capsule has an emotional category (feature) that corresponds to the text input. conclusively, the activation probability of these capsule must really be the highest when text sentiment fits the capsule characteristic, and the reconstructed features generated will correlate to the collected features. Furthermore, the training in this research has two objectives: the first, to enhance the activation probability which fits best the text sentiment whilst also reducing the error between both the text and reconstruction vector, and the another one, to reduce the capsule's activation probability while optimizing feature analysis error for feature vectors. As a result, the hinge loss function is employed to support this objective.

4.3 Learning Process of Proposed Network

For each Sentence matrix or word embedding encoding the text of the training batch; the categorization capsule with the highest activation probability transforms the sentiment matrix linearly, divides it into three linear vectors, and projects it to n subspaces. Then, for each subspace, compute the attention value to obtain head, and linearly transform to get Multi head, which connects to the sentence matrix and yields the output matrix. To produce characteristic matrices, apply convolution on the output matrix and bidirectional GRU long-distance encoding on the output matrix. The text entity feature vector is then obtained by performing global average pooling on it. Finally, run the capsule procedure to get the capsule's activation probabilities and the reconstruction features vector.

5 Experiment and Evaluation

We conduct research on an English datasets available on the Amazon public customer review data repository. The dataset include enriched data from year 2018 and records 12805 entries of user reviews on the software based products. The aforementioned dataset has been frequently utilized in sentiment classification tasks, allowing for a more accurate evaluation of the experimental findings. Each review text in the dataset possesses the user review entry in the range of 1 to 5, noting 1 is lowest and 5 being highest rating. Based on the proposed classification model, labels have been created using replacing 1 to 3 rating to "Bad" and ratings 4, 5 being "Good". Finally the labels for the data points are ready for binary classification using capsule network. The dataset was balanced and hence was ready to go for the preprocessing steps. The aforementioned preprocessing steps lead to the words embeddings which are ready to feed the proposed neural network.

Figure 11: Statistics for Datasets

Dataset	Train	Validation	Classes
Amazon Software Products	10.8K	1.9k	2

Tensorflow-keras is used in this research. The word embedding vector in the data set is initialized with a 1000-dimensional TFIDF vectorizor. We selected a set of widely used values for hyper-parameters based on prior research (Zhang et al.; 2017). Furthermore, in bidirectional GRU, the "Relu" activation function is employed, with a drop out rate of 0.25. We utilize Adam as an optimizer, with a learning rate of 0.001. Formally, several parameters of our approach will be tweaked slightly depending on the dataset. Table below lists the settings that were specified for each dataset. Finally, fifty epochs are used to achieve the final categorization result.

Figure 12: Network Parameters

Parameters	Value
Epoch	50
Learning Rate	0.01
Batch Size	256
Routings	3
Activation	Sigmoid
Number of Capsules	16
Embediing Vector length	1000

The following are the findings of the suggested model utilizing Capsules on the provided dataset. The baseline models in this article are based on a standard machine learning approach that was applied to the Amazon review dataset. We'll compare the outcomes using two distinct parameters called accuracy and recall. The baseline model implemented logistic regression, random forest and bernoulli naive bayes algorithms to perform the sentiment analysis (Shah et al.; 2021) and results are depicted below.

Figure 13: Evaluation of Classification Model

Model	Recall	Accuracy
Bernoulli naive bayes	0.95	93.17
Logistic regression	0.98	90.88
BI-GRU-Capsule network	1.00	70.47

5.1 Discussion

The proposed model seeks for the enhancement scope in sentiment analysis for amazon software products customer reviews over the traditional work carried out especially for this dataset. The evaluation matrices used for comparision are accuracy and recall defined earlier in the report. From the experimentation, several observation have been made. We have tried two vectorization methods namely TFIDF vectorizer and another by implementing word2vect model. The results were convincing in terms of evaluation matrices by using TFIDF vectorizer. The word2vect model had over generalize the context of text over its pre-trained learning. The model training was carried out on the aforementioned hyper-parameters for 50 epochs. after training the data of around 12805 recorded reviews, the accuracy achieved was 70.47 percent and recall of 1.

The accuracy can be interpreted the total number of correct predictions over the total predictions made. This resembles that the model posses satisfactory prediction capability. Though it is less compared to the base line model trained on logistic regression and Naive Bayes algorithm, this can be justified because of less training data.

The second and the most important evaluation matrix, we focused on the is recall. It is extent of optimism which model deliver meaning it shows how many correct predictions for positives are made out of total actual positives. This score outperformed the base model matrices for logistic regression and Naive Bayes algorithm. The recall value for proposed network is 1. This concludes that the total total number of true positive are correctly predicted with zero wrong predictions. Our objective is to get the correct sentiment out of the user reviews and we are actually being successful in doing so by achieving recall of 1.

Overall, the proposed model is capable of predicting the positive labels correctly to full extent and this work can be considered if the objective are solely based to reach out to customers which are actually satisfied with the products and make the promotions for the same.

6 Conclusion and Future Work

For text sentiment binary classification problems for software products review data, this research describes a capsule network combined with a CNN with additional bidirectional GRU implemented in it. The CNN will retriev the contextual local characteristics in the reviews to decrease gradient loss, according to an experimental evaluation of the model on this dataset. The bidirectional GRU runs through the whole review text to retrieve sentimental features, and produces the feature matrix of various layers of the review text upon global average pooling. Furthermore, this technique does not need the blending of language information.

In the future, research can be further improved in the scope of internal mechanisms of the emotional capsule, such as the attention mechanism; simultaneously, the larger dataset of same context can be utilised in order to get the better generalization while implementation of capsule networks. This will automatically will improve the accuracy matrix. Further one can think about improving the feature enhancing capacity, so the feature matrices can best serve emotional features and will help to think about improving the mode's effectiveness.

7 Acknowledgement

I would like to show my gratitude to Prof. Christian Horn, my supervisor, has been a great help and inspiration to me during this semester. I'd want to thank my friends and my family for their unwavering faith in me.

8 References

References

Athanasiou, V. and Maragoudakis, M. (2017). A novel, gradient boosting framework for sentiment analysis in languages where nlp resources are not plentiful: A case study for modern greek, *Algorithms* **10**(1).
URL: *https://www.mdpi.com/1999-4893/10/1/34*

Bhavitha, B. K., Rodrigues, A. P. and Chiplunkar, N. N. (2017). Comparative study of machine learning techniques in sentimental analysis, pp. 216–221, url : 10.1109/ICICCT.2017.7975191.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, KDD '16, Association for Computing Machinery, New York, NY, USA, p. 785–794.
URL: *https://doi.org/10.1145/2939672.2939785*

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H. and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation, *CoRR* **abs/1406.1078**.
URL: *http://arxiv.org/abs/1406.1078*

Das, B. and Chakraborty, S. (2018). An improved text sentiment classification model using TF-IDF and next word negation, *CoRR* **abs/1806.06407**.
URL: *http://arxiv.org/abs/1806.06407*

Elmurngi, E. and Gherbi, A. (2017). An empirical study on detecting fake reviews using machine learning techniques, *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, pp. 107–114,doi : 10.1109/INTECH.2017.8102442.

Hinton, G., Krizhevsky, A. and Wang, S. (2011). Transforming auto-encoders, Vol. 6791, pp. 44–51, doi : 10.1007/978–3–642–21735–7_6.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory, *Neural Computation* **9**(8): 1735–1780.
URL: *https://doi.org/10.1162/neco.1997.9.8.1735*

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features, *Proceedings of the 10th European Conference on Machine Learning*, ECML'98, Springer-Verlag, Berlin, Heidelberg, p. 137–142.
URL: *https://doi.org/10.1007/BFb0026683*

Kalchbrenner, N., Grefenstette, E. and Blunsom, P. (2014). A convolutional neural network for modelling sentences, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Baltimore, Maryland, pp. 655–665.
URL: *https://aclanthology.org/P14-1062*

Khurana, D., Koli, A., Khatter, K. and Singh, S. (2017). Natural language processing: State of the art, current trends and challenges, *CoRR* **abs/1708.05148**.
URL: *http://arxiv.org/abs/1708.05148*

Lilleberg, J., Zhu, Y. and Zhang, Y. (2015). Support vector machines and word2vec for text classification with semantic features, *2015 IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC)*, pp. 136–140, doi : 10.1109/ICCI–CC.2015.7259377.

Lin, A., Li, J. and Ma, Z. (2019). On learning and learned data representation by capsule networks, *IEEE Access* **7**: 50808–50822.
URL: *http://dx.doi.org/10.1109/ACCESS.2019.2911622*

Liu, B. (2015). *Sentence Subjectivity and Sentiment Classification*, Cambridge University Press, pp. 70–89, doi :10.1017/CBO9781139084789.005.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013). Distributed representations of words and phrases and their compositionality, *CoRR* **abs/1310.4546**.
URL: *http://arxiv.org/abs/1310.4546*

Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Barcelona, Spain, pp. 412–418.
URL: *https://aclanthology.org/W04-3253*

Pang, B., Lee, L. and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques, *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, Association for Computational Linguistics, pp. 79–86.
URL: *https://aclanthology.org/W02-1011*

Qian, Q., Tian, B., Huang, M., Liu, Y., Zhu, X. and Zhu, X. (2015a). Learning tag embeddings and tag-specific composition functions in recursive neural network, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Beijing, China, pp. 1365–1374.
URL: *https://aclanthology.org/P15-1132*

Qian, Q., Tian, B., Huang, M., Liu, Y., Zhu, X. and Zhu, X. (2015b). Learning tag embeddings and tag-specific composition functions in recursive neural network, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Beijing, China, pp. 1365–1374.
URL: *https://aclanthology.org/P15-1132*

Ramadhan, W., Astri Novianty, S. and Casi Setianingsih, S. (2017). Sentiment analysis using multinomial logistic regression, *2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*, pp. 46–49, doi : 10.1109/ICCEREC.2017.8226700.

Ren, H., Su, J. and Lu, H. (2019). Evaluating generalization ability of convolutional neural networks and capsule networks for image classification via top-2 classification, *CoRR* **abs/1901.10112**.
URL: *http://arxiv.org/abs/1901.10112*

Sabour, S., Frosst, N. and Hinton, G. E. (2017). Dynamic routing between capsules, *CoRR* **abs/1710.09829**.
URL: *http://arxiv.org/abs/1710.09829*

Saqib, M.S., Kundi, F.M. and S., A. (2018). Unsupervised learning method for sorting positive and negative reviews using lsi (latent semantic indexing) with automatic generated queries, Vol. 18.
URL: *http://paper.ijcsns.org/07$_b$ook/*201801/20180107*.pdf*

Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N. and Shah, N. K. (2021). Sentiments detection for amazon product review, *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, doi : 10.1109/ICCCI50826.2021.9402414.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Seattle, Washington, USA, pp. 1631–1642.
URL: *https://aclanthology.org/D13-1170*

Tai, K. S., Socher, R. and Manning, C. D. (2015a). Improved semantic representations from tree-structured long short-term memory networks, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Beijing, China, pp. 1556–1566.
URL: *https://aclanthology.org/P15-1150*

Tai, K. S., Socher, R. and Manning, C. D. (2015b). Improved semantic representations from tree-structured long short-term memory networks, *CoRR* **abs/1503.00075**.
URL: *http://arxiv.org/abs/1503.00075*

Tang, D., Qin, B. and Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Lisbon, Portugal, pp. 1422–1432.
URL: *https://aclanthology.org/D15-1167*

Xu, J., Xu, R., Lu, Q. and Wang, X. (2012). Coarse-to-fine sentence-level emotion classification based on the intra-sentence features and sentential context, *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, Association for Computing Machinery, New York, NY, USA, p. 2455–2458.
URL: *https://doi.org/10.1145/2396761.2398665*

Zhang, H., Wang, J., Zhang, J. and Zhang, X. (2017). Ynu-hpcc at semeval 2017 task 4: Using a multi-channel cnn-lstm model for sentiment classification, pp. 796–801, doi : 10.18653/v1/S17–2134.

Zhang, L. and Liu, B. (2017). Sentiment analysis and opinion mining, *in* C. Sammut and G. I. Webb (eds), *Encyclopedia of Machine Learning and Data Mining*, Springer US, Boston, MA, pp. 1152–1161.
URL: *https://doi.org/10.1007/978-1-4899-7687-1*$_9$07

Zhao, W., Guan, Z., Chen, L., He, X., Cai, D., Wang, B. and Wang, Q. (2018). Weakly-supervised deep embedding for product review sentiment analysis, *IEEE Transactions on Knowledge and Data Engineering* **30**(1): 185–197, doi : 10.1109/TKDE.2017.2756658.