2020-08

SPEECH ACT AND INTENT CLASSIFICATION USING A CONVOLUTIONAL NEURAL NETWORK

ATLAW, MULAT

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF COMPUTING

SPEECH ACT AND INTENT CLASSIFICATION USING A CONVOLUTIONAL NEURAL NETWORK

BY

ATLAW MULAT

BAHIR DAR, ETHIOPIA

August, 2020

SPEECH ACT AND INTENT CLASSIFICATION USING A CONVOLUTIONAL NEURAL NETWORK

BY

ATLAW MULAT

A Thesis Submitted to the School of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in Partial Fulfillment for the Degree of Master of Science in Computer Science in the Faculty of Computing

ADVISOR: Mr. Seffi Gebeyehu. (Ass, Prof)

BAHIRDAR, ETHIOPIA

August, 2020

DECLARATION

I, the undersigned, declare that the thesis comprises my own work. In compliance with internationally accepted practices, I have acknowledged and refereed all materials used in this work. I understand that non-adherence to the principles of academic honesty and integrity, misrepresentation/ fabrication of any idea/data/fact/source will constitute sufficient ground for disciplinary action by the University and can also evoke penal action from the sources which have not been properly cited or acknowledged.

Name of the student: Atlaw Mulat Signature: _____

Date of submission: _____

Place: Bahir Dar

This thesis has been submitted for examination with my approval as a university advisor.

Advisor Name: Mr. Seffi Gebeyehu (Ass, Prof)

Advisor's Signature: 

Bahir Dar Institute of Technology-Bahir Dar University

School of Research and Graduate Studies

Faculty of computing

THESIS APPROVAL SHEET

Name of Student: Atlaw Mulat Signature Date 14/12/12 E.C

As members of the board of examiners, we examined these thesis entitled Speech Act and Intent Classification using Convolutional Neural Network by Atlaw Mulat. We hereby certify that the thesis is accepted for fulfilling the requirement for the award of the degree of Masters of Science in Computer Science

Approved By:

Name of Advisor:	Signature	Date
Seffi Gebeyehu (Ass. Prof)		August 20/2020
Name of External Examiner:	Signature	Date
Wondwossen Mulugeta (PhD)		
Name of Internal Examiner:	Signature	Date
Abinew Ali		20/08/2020
Name of Chairperson:	Signature	Date
Mekonen Wagaw (PhD)		14/12/2012 E.C
Name of Chair Holder:	Signature	Date
Haileyesus Amsaya		14/12/12 E.C
Name of Faculty Dean:	Signature	Date
Belete Biazen		14/12/12 E.C

ACKNOWLEDGEMENT

At the outset and foremost, I would like to praise my almighty GOD, who favors me from beginning to end of this study, Without GOD noting happen forever.

I would like to express my deepest gratitude to my advisor, Mr. Seffi Gebeyehu. (Ass, Prof), for his continuous support and guidance throughout the various stages of this thesis. He provided critical and useful feedback and suggestions on how to address the research problems systematically and tactfully. His punctuality, encouragement and on time contact has always inspired me and help to overcome good result and the preceding of work in time. As my supervisors, his observations, guidance and comments help me to find the right direction of the research and to move forward, and to complete the thesis.

ABSTRACT

The spoken language understanding is an emerging field between speech processing and language understanding from the human spoken utterance. It can apply in different sectors for representing the meaning of a human spoken utterance in different domains. The analysis and identification of the speech act and intent in the use of language for the speaker motivation in political speech candidates are important factors for the willingness of the speaker and the audience. But there is no an annotated speech act and intent recognition corpus for the Amharic language from the spoken utterance, due to these, the politician transmits never known what is say and what is meant in the transmission and the audience also not creates a common understanding of the transmitted information, because of the different interpretations of the utterance. In this study to tackle the problems, we have been proposed a convolutional neural network approach with a word and sentence embedding technique for analyzing and identifying a once utterance speech act and intention in political speech candidate. We have been used a word2vec approach for converting each word in the corpus for vector values with a continuous bag of word architecture and also used a mean embedding approach for converting each utterance value in numeric value representation. We have applied a convolutional neural network approach, for extracting deeply features from the distributional matrix representation of each speech utterance and also for the identification of each speech act and intention utterance class by using the different activation functions. We have achieved an accuracy of 92.5%, 89.3%, and 83.8% for speech act, speech intent and the intent based speech act classification as Softsign function of SoftMax classifier, and also 88.1% and 56.8% for speech act and the intent based speech act classification as ReLu function and 59.6% for speech intent as Softplus function for sigmoid function in the output layer respectively.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ASR	Automatic speech recognition
BiGRU	Bidirectional Gate Recurrent Unit
CBOW	Continuous Bag of Word
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DM	Dialogue Management
ELMo	Embedding from Language Model
GloVe	Global Vectors for Word Representation
ML	Machine Learning
NLP	Natural Language Processing
NLU	Natural Language Understanding
RAM	Random Access Memory
RBWE	Ranking Based Word Embedding
ReLu	Rectified linear Unit
SDS	Spoken dialog system
SLU	Spoken Language Understanding

LIST OF FIGURE

LIST OF TABLE

CHAPTER ONE

1. INTRODUCTION

The spoken dialogue system is a computer system that human ability to communicate with the computer. In the recent era, the SDS has been applied in different situations for tackling the traditional activity of humans to transfer technology-based activity as a human can communicate with the robot, a robot with a robot, and also any activity that accepts voice at the same time give a voice response (Ramón, Zoraida, & et al, 2014). To perform this activities it can follow a complex structure of five components. One of them is spoken language understanding. The Spoken language understanding (SLU) is an emerging field in between speech and language processing, for investigating human to machine and human to human communication by leveraging technologies from signal processing, pattern recognition, machine learning, and artificial intelligence (Gokhan & Renato, 2011)**.** The SLU systems are designed to extract meaning from the speech utterance and its application is vast, from voice search in mobile devices to meeting summarization, domain detection, intent determination, and slot filling. The spoken language can be translated into its corresponding plain text by the application of an automatic speech recognition task. The plain text can describe the speaker's intended or wanted by applying a natural language processing (NLP) technique integrate with other artificial intelligence tool or machine learning approach for knowing the use of language in particular situations are lead to the pragmatic of utterance (Renato , 2013).

In recently, SLUs were applied in different sectors from the numerous wise use of applications like in educational society, health care, political sector, public service, and also a business area for differentiating the speech act and intent in wisely format. The speech act can determine the function of utterance. The speech act was first developed by (Austin , 1962) to explain an utterance within a natural language to get feedback and also performed by a particular word often depends on the speaker intention and the context in which the word uttered (Kent , 2018)**.** The theory of (Austin , 1962) stated that, there is three class of speech act for representing the speaker utterance such as locutionary, illocutionary, and perlocutionary in political sectors. The locutionary act is the act of saying something, illocutionary act is the successful realization of the

speaker intention, which is a product of the listener interpretation, and the perlocutionary act is the effect or influence on the feelings, thoughts or actions of the listener or hearer.

In the Illocutionary speech act (Searle, 2012) stated that there is five class of speech act for transmitted information such as assertive, commissive, directive, expressive, and declarative. Assertive means to commit speaker to the truth of the expressed proposition, like the utterance are stating, claiming, reporting, announcing, suggesting, putting forward, boasting, and concluding. Directive means statement that compel or make another person action fit the propositional element, causing the hearer to take a particular action, request, command or advice, asking, ordering, inviting, and begging. Commissive means to commitment speaker to some future actions, like the utterance as a promising, offering, swearing, planning, vowing, betting, and opposing to do something. Expressive means to count as the expression of some psychological state, like the utterance are thanking, apologizing, congratulating. Declarative means the statement are used to say something and make it so, such as pronouncing someone guilty, resigning, dismissing, accepting, declaring a war. The researchers are mainly focused on the meaning of utterances performed by a speaker and/or interpreted by a hearer. The intention represent the speaker original purpose of saying the utterance. Once in the political sector can be applied for differentiating the speech act and intention of the speaker, by extracting meaning from the utterance are advantageous for current and future situation.

However, to the best of our knowledge, in the Amharic language for representing the meaning of human spoken utterance are advantageous for recognizing the function of the utterance, identifying the domain of the utterance intention, interpret and analyze the explicit and implicit transmitted information from the speaker and also the undertaking or initiation of information for the audience by integrating other discipline technology like machine learning, deep learning or artificial intelligence. They can be achieved a better performance by applying a convolutional neural network on the pertained word embedding by handling the out of vocabulary word and sparse data set exists for multi-label speech act classification (Kyoungman & Youngjoong, 2018). They can be applied a bidirectional Gated Recurrent Unit as a classification purpose for representing the speech act and target party at the campaign communication for the voters can be making a good decision for the election party (Shivashankar, Trevor, & Timothy , 2019).

In political sectors, once the election manifesto announce can influence a party's reputation, credibility, and competence, which are primary factors in voter decision making (Pablo , 2014) and also election manifesto have been analyzed for prospective and retrospective message (Stefan, 2018). As coming to the Amharic political speech, the transmitted message and discussion in political speech come from the political party debate, political party manifesto, parliamentary meeting, and also conversations from studio program are hugely based on the ideology of the speaker intention. As the transmitted information stays in a necessary path for present and future direction, but the variability of discussion transmitted message have been changed your direction, they have also create a mismatch ideology between the speaker and community, and also people appeal to a negative conflict emotion. This situations come from the politicians are unaware of the fact that there is a link between what is said, what is meant, and the action conveyed by what is said. The community cannot also create a common understanding of the speaker propaganda and in what way the transmitted message from the speaker has been a failure and/or success. This situation come from the community people have a different interpretation that can influence the success of the politician ideology.

They can also require an annotate speech act for each utterance of generalize ideology and also the intention behind the transmitted information that the users will be belief or desire or intend of the speaker, for analyzing and identifying of once personal stating ideology for researchers, politicians, the audience in the meeting. In this study, we have implemented a speech act and intent classification for Amharic political speech, and also mainly concerns with the three illocutionary speech acts of intent recognition (assertive, commissive, and directive). We have used a word2vec model of word embedding techniques for representing the word features into vector space representations, sentence embedding techniques for extracting meaning from the utterance, finally, we have used a convolutional neural network (CNN) deep learning approach for feature extraction from the matrix representation of utterance for achieving a better speech act and intent classification in Amharic political speech.

1.1. Statement of the Problem

The Amharic language has been one of the working language in Ethiopia. Most of the researchers are foremost concentrated on the Amharic language for analyzing natural language understanding tasks in different sectors for integrating other technology-based activities. For analyzing and

representing the semantic meaning of utterance have been applied in some sectors as for identifying the overall intention of the user speech act utterance to provide a better user oriented service, handling the flow of message, and also compromise, understand and utilize the communication between the speaker and audience. As main concerns in the political sector for representing the semantic features of each speaker utterance.

Once the political party can campaign the election policy to the voters or community, they can transmit and/or address in a different format. While we have taken the human spoken utterance for representing each utterance meaning are favorable to recognize the function of the utterance and identifying the domain of the utterance intention in current and future direction is an advantageous situation for creating political, social, and economic relations among the community. Moreover, in the Amharic language, there is no any attempt for analysis the function of utterance and identification of its intention from the human spoken utterance during transmitted information to the concerned body, due to this the politicians, audience, and other researchers cannot easily analyze the transmitted information from such candidates for the identification of its speech act and intent.

As stated in related work, the speech act and intent analysis and identification have been performed separately in different situations. As for classifying each speech act class depends (in hierarchical format) of its intentions are more favorable for creating a general or an accurate ideology. As for estimating the classification performance of each speech act and intent and also combine both of the intent and act the activation function has a great role in the proposed architecture model.

Besides, as described in related work and as per the knowledge of the researcher, no research has been done for the Amharic language. To this end, this research have been answered the following question:

- How to develop an Amharic speech act and intent recognition system?
- Which activation function is suitable for speech act and intent classification from textual data?
- To what extent the identification of speech act and intent from textual data will be achievable?

1.2. Objective of the Study

1.2.1. General Objective

The general objective of this study is to develop a speech act and intent classification for Amharic political speech using a convolutional neural network approach.

1.2.2. Specific Objectives

To achieve the general objective of this research, we have considered the following activity:

- To gather and analysis the features of speech act and intent for the development of Amharic political speech.
- To design the model of speech act and intent classification for Amharic political speech.
- To identify a suitable activation function for the Amharic political speech act and intent classification system.
- To investigate the effect of activation function for the Amharic political speech classification.
- To evaluate the performance of the proposed model for Amharic political speech.

1.3. Methodology

In order to accomplish this research, we have considered data gathering and also prepare the data with its class, selection of appropriate development models and approach, and testing methodology for evaluating the proposed approach fit with the model. We have been discussed each methodology as follow as:

1.3.1. Data Collection

In data gathering we have to consider two things these are where are the data source with the type of data and the preparation of the collected data with its class based on the characteristics features. In this study, for Amharic political speech there is no well-defined corpus available for grasps each speech with respective classes, because of this we have collected the data set from the politicians stating its own idea to social media and political party manifesto.

As describes above, we have used a primary data source of archive techniques from document file, and online available plain text format. We have labels each utterance with the respective class with the illocutionary force of verbs within the utterance. Like assertive for expressing the utterance reporting, claiming, stating for some situation. The commissive for commitments taking future actions of utterance promise, offering to do something. The directive orders to take a particular action in the utterance requesting, ordering, advice. We have applied a text cleaning for remove unwanted script characters, numbers or punctuation marks from the utterance, and also removal of stop word (can occur frequently in the documents that do not affect the classification accuracy). We have used a word tokenization technique for converting each text document into word samples for representing each word into a corresponding vector representation.

1.3.2. Research Design

In this research, we have been followed by a scientific research procedure. Research design is a set of procedure or method that effectively address the problems, guidelines for using an appropriate framework in the research flow chart. The goal of this research is to develop a classification model as an experimental research design approach have been conducted. As an experimental research method have been employed to solve the problem in a wisely manner.

1.3.3. Model Design

One of the reasons that SLUs are a difficult problem to solve is the fact that, unlike human beings, computers can only understand number format representation from any object. In textual data classification, there is applying a machine learning, deep learning approaches integrate with the word embedding technique for achieving efficient ranking result. We have select a word embedding technique for representing the plain text into vector format for extracting features from each word, a sentence embedding technique has been applied for representing the utterance into vector representation from the word embedding result. Finally, we have been applied a CNN deep learning approach for extracting features from the sentence embedding result table for each utterance and also classifying each utterance value.

1.3.4. Tool and Implementation

In order to implement the theoretical part into a real world view, as an editor, we have used a PyCharm community editor IDE with the python programming language. We used a python 3.7 for invoking and accessing additional libraries and modules as want to use and also it is simple to encoding logic programming concept to view the user aspect. The PyCharm editor is also better integrated with the python library from the anaconda framework. The tool for preprocessing the word embedding data feature extraction, training the model and classification purpose, we have select a Keras (which is an application programmer interface on top of the Tensorflow for deeply extracting and training the model for better classification result.

1.3.5. Evaluation Procedure

In this thesis, for classifying the speech act and intent from the utterance, as working in an actual manner and also give some trust to the prediction model. We have followed four-phase passed. In the first phase, we have prepared and analyzed the raw data sets to remove unwanted scripts, numbers, and texts from the utterance. In the second phase, we have applied a word embedding (word2vec) technique for representing each unique word into a vector representation. In the third phase, we have applied a sentence embedding technique for converting each utterance in to vector representation format. Finally, we have used a holdout technique for dividing the data set into training to train the model and testing data set to test the model, and also feeding the matrix representation of each utterance value to the CNN for extracting a feature from the utterance, to train and test the model and also applied a different activation function for classifying correct values.

1.4. Scope of the Study

In SLU perspective situations for knowing the speech act of the transmitted message from the speaker and the intent of the utterance are more favorable for investigating a good relationship between the speaker and the audience. The speech act in political speech categorize into five classes these are assertive, directive, expressive, declarative, and commissive for generalizing the intent of the speaker utterance. Furthermore, in each speech act class, there is a subcategory for example in commissive such as promise, offering, swearing, etc. for expressing once intentions within the speaker utterance in detail expression. In this research study, besides the other researchers gives the most percentage values for each class, and also due to the data available for

most concerns of the Amharic political speech candidates. We have been considered the three main illocutionary act category classes like assertive speech act with its intent type report and claim, as in directive speech act with its intent type request, order and advice and also in commissive speech act with its intent type promise and offer without considering the intention extractions phase of the politician utterance and also identifying the target point of from the transmitted information of politician speech utterance.

1.5. Significance of the Study

After the completion of this research, on the speech act and intent classification in political speech has been some contribution as researchers and also give some benefits to the society, the audience in the meeting, individuals or politicians for analysis and classifying once own political speech ideology. Some of these are:

- We have analyzed and classify political speech acts and also represent the intent of the speech act for once a political candidate.
- To help the politician have identified the current role of circumstance and the feature goal and intent for creating a good awareness for the audience.
- To help the politician have also analyzed the socio-political and economic contexts of Ethiopian before, present, and futures directions at the times of delivering its own speech easy to comprehend the message.
- To help the audience have created a common understanding of the concerned politician speeches within an effective and efficient time manner.
- To help the audience have to understand the success or failure of some types of politician speeches or how breakdowns in communication can occur.

1.6. Organization of the thesis

This thesis paper organized as follows. In chapter two, we have been discussed the SDS and overview SLU in different sectors, the proposed deep learning approaches in deeply and also related works. In chapter three, we have been discussed the proposed word embedding CNN based speech act and intent classification models in detail. In chapter four, we have been analyzing the data sets used and experimental results with the selected activation function. In

chapter five, we have been discussed the conclusion, contribution, and recommendations in briefly.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Introduction

Spoken dialog system (SDS) is a computer system that allows to be communicate or convince with human with voice. Recently, it is a new breed of interfaces that enable human to communicate with machine naturally and efficiently using a conversational ecosystem (Victor & Stephanie, 2008). The SDSs are built by integrating several independent components, which are taking the human speech signal, processing the signal within the entire component and also finally give the response to the human.

The foundation of SDS is complex to set up because the implementation requires employing a number of components to process the human language independently, which is a very complex task. The first component is, Automatic speech recognition (ASR), this module is to receive the user speech and generate as output a recognition hypothesis, which is the sequence of words that most likely corresponds to what the user has said (Ramón, Zoraida, & et al, 2014). The second component is Spoken language understanding (SLU), this module is to obtain a semantic representation of the input, which typically is stored in the form of one or more frames. The third component is Dialogue Management (DM) can accept the output of the SLU, this module is to decide what the system must do next in response to the user input (Ramón, Zoraida, & et al, 2014). The fourth component is natural language generation can accept the output of the DM decision to transform into one or more utterance in text format that either must be grammatically and semantically correct or checking the coherence with the current status of the dialogue (Lemon, 2011; Víctor, Eduardo , & et al, 2012). The final component is text to speech synthesizer, this module can accept the utterance text to transform corresponding speech signals that the human can response.

2.2. Overview of Spoken Language Understanding

The spoken language understanding (SLU) is currently an emerging field in the intersection of speech processing and natural language processing (NLP) by leveraging technologies from machine learning (ML), deep learning and artificial intelligence (AI) (Gokhan & Renato, 2011).

In recently, SLU can be an emerging technology between the human to machine interaction, and also human to human interaction. The SLU systems have used a text-based natural language understanding (NLU) approaches for processing a sequence of word hypotheses generated by the ASR for extracting meaning from the natural language. Sentence of a natural language are sequences of words belonging to a word lexicon. Words of a sentence have associated with one or more data concept this leads to the meaning of the sentence. These words can be selected and composed to form the meaning of the sentence.

In the SLU, for identifying and analysis the words exist in the utterance have been useful for representing the semantic meaning of utterance. It can also automated the human to machine or human to human interaction tasks as technology-based, from these the SLU have becomes far from unreachable dreams as numbers of tasks can be exist. (Gokhan & Renato, 2011). The task for technology-based in the area of SLU is intent determination, spoken utterance classification, voice search, and also other activities. The intent determination is simply what is the intended or wish of the transmitted and/or response message. The spoken utterance classification is simply the speech act that once to be convinced or the combination of several acts at once for distinguishing once speaker to the others. The voice search for extracting query from the database by actively investigated speech understanding technology, spoken question answering and also extracting the correct semantic frame from the spoken utterance.

In the SLUs, these and other tasks could exist for automating in present and for future in a different ecosystem like in educational sectors for determining the student to teacher interaction for creating smooth coordination and cooperation in the learning-teaching process like for giving clarification, question, request, command, promise. In the health care sector for creating an automated verbal communication for sharing the intention between the patients and the specialist in the hospital (Yue, Xinyu, Shuhong, & et al, 2017). In the public service sectors for creating a good fairing environment between the request and response. In the political sectors for knowing the message of transmitted information from the speaker and also the listener. In the business sectors for the seller knows the intended of the customers form the workshops to buy from the speech and also the customer knowing the wish of the seller man in the workshops. In religious sectors, and also other sectors are advantageous.

2.3. Political Speech

Amharic has been one of the Semitic languages spoken in north-central Ethiopia. Next to Arabic, it is the second most-spoken Semitic language in the world and it is one of the working languages of the Federal Democratic Republic of Ethiopia. It has been the second-largest language in Ethiopia (after Afan Oromo, a Cushitic language) and possibly one of the five largest languages on the African continent (Demeke, 2010). As a result, it has an official status and used nationwide. Despite it has a large speaker population, the language has little computational linguistic resources (Atelach & Lars, 2007).

Speech is the communication or expression of thoughts in spoken words. It is used to produce the words to signal format to be hired or listen by others for creating well-being communications (Faiz, 2017). The political speech is an expression or communication that can be taken or acted by the government organizations for resolving something new for the concerned community and also converges as a whole ideology. The political language deals with the use of power to organize people mind and opinion comes to once an ideology (Suhair M. , 2015). It has been an instrument used to control society by convincing, deceive, cheating, or some trade-off ideas that can be realized. In political speech, the ideas or ideologies need to be conveyed through language so that they are agreed upon by the receivers as well as by others who may read or hear parts of the speech afterward in the media. Once the politician speaks, they have used words within the utterance as a means of establishing and maintaining social relationships, expressing feelings, and selling ideas, policies, and planning program in any society.

In the political speech, for predicting the act of one politician utterance depends on the word exist in the utterance. In the extracting of meaning from the utterance, they have been following either semantic or pragmatics direction. The semantic is a conventional meaning conveyed by the use of words, phrases, and sentences of a language. The pragmatic is to interpret the meaning and function of words in different speech situations. The pragmatics is also study the relationship between language and its social context in the process of communication between the speaker and hearer (Suhair M. , 2015). It has been mainly focused on the meaning of utterances performed by a speaker and/or interpreted by a hearer, this pragmatic studies in political sectors can lead to speech act. The speech act is a linguistic action to transmit a message to the hearer, this has been giving the intention of the speaker. The intention represent the speaker original

purpose of saying the utterance. In the political speech situation for extracting the intent from the speaker utterance are an advantageous for compromise the belief or desire or wanted behind the transmission information to the audience (Otis, 1969).

2.3.1. Speech Act

In political speech, the speaker and audience have been communicating, understanding, and creating a harmonious relationship about the transmitted information of its target point of view by invoking some intermediate technology. From these speech is premised on the fact that people perform various actions through the use of words and when utterances are made. In the utterance, there is performs a particular act that refers to the speech act of the transmission message. The speech act is a primitive abstraction or an approximate representation of the illocutionary force of an utterance (Otis, 1969).

According to (Austin , 1962), the speech act has been categorized into three classes for expressing the utterance. These are locutionary, illocutionary, and perlocutionary acts. The locutionary act is the act of saying something or the act of producing an utterance. The illocutionary act is identified by the explicit performative verb. It is simply what one does in saying it or the successful realization of the speaker intention. The perlocutionary act is the effect or influence on the feelings, thoughts, or actions of the listener/hearer. It is simply what one does by saying the speaker.

According to (Searle, 2012), as a convention, the speech act has been classified based on the illocutionary force, by identifying the familiar verbs within the utterance, as they can express the speaker intention and also a product of the hearer interpretations. From these, it has been classified the illocutionary act into five class such as assertive, commissive, directive, expressive and declarative speech act type.

The assertiveness is to commit the speaker to be the truth of the expressed proposition. As the utterance can illustrate a stating, claiming, reporting, answering. The audience has been accepting or argue the transmitted proposition can be a truthfulness values.

E.g. **እያካሄደው ባለው በጥልቀት የመታደስ እንቅስቃሴ በሃገራችን ኣዲስ ክስተት እየተፈጠረ ይገኛል** => as the speaker to express the truth stated information to be performed as served for the audience as reporting. The audience simply accept the expressed proposition.

The commissiveness is committing the speaker to some future actions. As the utterance illustrates a promising, offering, swearing for to do something. As the audience can also belief or desire as doing or something come up in the future in the forwarding conversation.

E.g. ***እኛ ለጋስ እና ሞቃታማ ሀገር እንሆናለን*** => as the speaker to realize or give a hope ideology for futures to do something as changing the life standard conditions of the community.

The directive is a statement that compels or make another person's action fit the propositional element. It is usually used to give orders thereby causing the hearer to take a particular action, request, order, or advice. As the user can belief the speaker can give some instructions for that environment conditions.

E.g. ***የአባሎቹና አመራሮቹ እስር ከድርጊቱ ጋር የማይገናኝና ፖለቲካዊ መሆኑን በመግለጽም፤ አባላቱ በአስቸኳይና ያለ ምንም ቅድመ ሁኔታ እንዲፈቱለትም ስንል በትህትና እንጠይቃለን*** => as the speaker can transmit an instruction that will be take an actions.

2.3.2. Speech Intent

In political speech, the intention represents the speaker original purpose of saying the utterance or it is the aim, goal, or purpose of the utterance transmit a message to the audience. As the speaker uses in political sector for the hearer to be impressed, convince, and even to deceive for saying utterance (Otis, 1969). For knowing the intention of any available organization are advantageous like in telecommunication center, for identifying the customer call intent like sales, job seeker, services, or vendor to the service provider for reducing the engagement framework in the telecom center (Junmei & William, 2019). In the health sector, for the patient and the medical workers as identifying the intention conversion between each candidate like give direction, question, instruction, clarification or an acknowledgment (Yue, Xinyu, Shuhong, & et al, 2017). In educational sector for facilitating the learning teaching process between student and teachers as like giving instruction, clarification, questions or any other suggestions.

The intent determination in political sector also have numerous advantage for present and future direction as putting some relief map for the coming generations as instructional, pride and also gaining some recognitions from the populations. In this study, we have consider seven intent

class in collected political speeches. Some examples for each class of the intention in political speech that can be adhere:

I. Report

Report is as the speaker to describe, disclose, announce, make known, issue about state, express, narrate and also advertise the transmitted information to the audience. On the other hand, the speaker can also communicate, divulge, explain, post, state, document, draft and precise the information that can be transmitted. In general, the user can be belief the speaker propositions are acceptable or be truth.

No	The word infers to be	As the utterance can be reflected
1	ማስተዋወቅ (Announce)	አንድን ተሰርቶ ወይም ተፈጽሞ በአጠቃላይ ለህዝቡ ወይም ለተፈለገው አካል በአደባባይ በይፋ እንዲታወቅ ማድረግ
2	ዝርዝር (Explanation)	የተሰሩ ስራወች ምን እንደሚመስሉ እና አመጣጣቸው ምን እንደሚመስል በሚገባ መልኩ ለተፈለገው አካል ማድረስ ፤ ስለተሰሩት ስራወች ትንታኔውያ መረጃ መግለጽ
3	በትክክል (Truthfulness)	የሚተላለፉት መረጃወች ሙሉ በሙሉ እውነታን መሰረት ያስደገፉ መሆናቸውን ለማህበረሰብ ማሳመን እና በሚጠየቁበት ጊዜ ጥቅም ላይ የሚውል መሆን አለባቸው።
4	መግባባት (Communicate)	የሚተላለፈው መረጃ ከማህበረሰቡ ጋር ጥሩ ግንኙነት የሚፈጥር መሆን በግልጽ ለህብረተሰቡ በሚገባ መልኩ ማስተላለፍ መሆን አለበት።

Table 2. 1 Report Guideline

E.g. ***እያካሄደው ባለው በጥልቀት የመታደስ እንቅስቃሴ በሃገራችን ላዲስ ክስተት እየተፈጠረ ይገኛል*** => is the proposition that the speaker to be transmitted (**P**). The audience can be belief that **P** to be truth as stated information serves as final report from the speaker intention type.

II. Claim

Claim is as the speaker can be protest, argue, defend, testify, objection, complaint, disapproval, disagreement, opposition, dissent for once transmitted information by other group members or politician party. On the other hand, they can declare, challenge, proclaim, adduce, give proof,

and evidence for the transmitted information and also maintain, asseverate, represent, uphold, certify, and affirm of the transmitted information. In general, the user can be belief the speaker can appeal to the transmitted propositions by others.

No	The word infers to be	As the utterance can be reflected
1	የይገባኛል ጥያቄ (Claim)	በሌሎች የተሰሩና የተፈጸሙ ነገሮችን ስለእውነታ መረጃቸው ያላቸውን ቅሬታ ወይም አቤቱታ ለሚመለከተው አካል መግለጽ ማለት ነው።
2	ተቃውሞ (Protest or Oppose)	በሌሎች ቡድን የተሰሩ፣ የመጡ ሀሳቦችን ሙሉ ለሙሉ አለመቀበል እና ለማህበረሰቡ ያቀረቡት ሀሳብ ውድቅ እንደሆነ መግለጸ ማለት ነው። የቀረቡት ነገሮች ሁሉ የተሳሳቱ ወይም አጥጋቢ አለመሆኑን መግለጸ።
3	ክርክር (Argue or Contend)	ከሌላ ቡድን የመጡ ሀሳቦችን አቅጣጫቸው፣ የወደፊት መፍትሄቸውና የያዙት ሀሳብ ምን እንደሚመስል የቅሬታ ሀሳብ ማቅረብ ማለት ነው።
4	ዋስትና (Guarantee or Warrant)	ለቆሙለት ማህበረሰብ፣ ለሚያቀርቡት መረጃና ለሚሰሩት ስራ ከሚሰነዘሩበት ወንጀልና መሰል ጥቃቶች የመከላከል ብሎም በህግ ከለላና ጥበቃ ስራ ማዋል መቻል ነው።

Table 2. 2 Claim Guideline

E.g. ***ላሁን ያለው ላመራር በያንዳንዱ ማሳና ስድራ ቤት ገብቶ ላርሶ ላደርን መደገፍ እንደዋነኛ የስራ ሃላፊነቱ መሆኑ የተገነዘበ ሃይል ላይመስለኝም*** => is the proposition that the speaker to be argue or protest that once idea comes from (**P**). The audience belief that the speaker can appeals to stated **P** as a claim intention type.

III. Promise

Promise is as the speaker can be pledge, swear, covenant, solemnly promise, and give hope of to the transmit information for the audience to do something in the future. On the other hand, the speaker can also give an undertaking, give an assurance, point to, denote, bespeak for giving information to others and also warrant, be emphatic, absorb and for tell about the transmitted information to the audience. In general, the user will be belief the speaker intend to the propositions of the transmitted information to do something in the future.

No	The word infers to	As the utterance can be reflected

	be	
1	ቃል ገባ (Pledge or Covenant)	ለወደፊት ለምንሰራው፣ ለምንፈጽመው ነገር ለቆምንለት ማህበረሰብ ተስፋ የጫረ ቃል መግባት ማለት ነው።
2	ተስፋን ስጥ (Give Hope)	ለወደፊቱ ለቆሙለት ማህበረሰብ የሆነ ነገር ተሰርቶ፣ ማናልባትም ያሰቡት ነገር በሆነ አጋጣሚ እንደሚከሰት በጥሩ ምክንያት ለህዝቡ በተስፋ ወይም በጉጉት መሙላት መቻል ማለት ነው።
3	ማበረታታት (Bespeak or Engage)	ለማህበረሰቡ የጀመሩትን ስራ ወይም አላም ለመፈጸም ይችሉ ዘንድ በራስ መተማመን እንዲኖራቸው ጥሩ ሀሳቦችን ማዝለቅ መቻል ነው።
4	እርግጠኛ (Signify or Give of an Assurance)	አሁን ላይ ለተሰሩና ወደፊት ለሚሰሩ ነገሮች እርግጠኝ የሆነ እወቀት ለማህበረሰቡ ማስረዳትና እንዲያውቁት ለማድረግ መጣር ያለው ፍላጎት ነው።

Table 2. 3 Promise Guideline

E.g. ***ሰራተኞቻችንን የሚጎዳ እና ነጻነታችንን የሚቀንስ ማንኛውንም የንግድ ስምምነት በጭራሽ ላለመፈረም ቃል እገባለሁ*** => is the proposition (**P**) that the speaker to be give a hopeful ideology realized to do something in the future. The user belief the speaker intend to **P** as a promising intention type.

IV. Offer

Offer is as the speaker can be provide, give, supply, donate, and contribute for something to the public. On the other hand, the speaker can also award, accord, tender, volunteer, propose, put forward, handle, make known, educe, deliver, proffer, and give an opportunity for the audience or the public that can gain something for you. In general, the user will be desire the speaker intend to the transmitted propositions for future actions as gain harmony relationship between you and the audience or community.

No	The word infers to be	As the utterance can be reflected
1	አስረክበ (Give or Submit)	ለተለያዩ መሰረታዊ ልማቶች እንደርዳታ መሰል ነገሮችንና መሰል አነቃቂ ሀሳቦችን ለማህበረሰቡ ለማስረከብ በቁርጥኝነት የመነሳት ተስፋ የመሰነቅ።
2	ልገሳ (Donate)	በሀገር፣ ማህበረሰብ፣ በተለያዩ የመንግስትና የግል ድርጅቶች ላይ

		የሚንጸባረቀውን ችግር ለመቀረፍ በተቻለ መልኩ ገንዘብ በመራዳት፣ የተለያዩ የእቃ መሳሪያወችን በመስጠትና በሙያዊ ብቃት በመሳተፉ መቻልን ማሳየት ነው።
3	አቀረብ (Provide or Supply)	የሚፈለግውን ነገር ለህዝቡ በተፈለገው ስአትና ቦታ ለተፈለገው ስራ የማቅረብን ችሎታን ለማህበረሰቡ ይፋውያ መግለጫ መስጠት መቻል ነው።
4	ርህራሄ (Tender)	ቆምንለት ላሉት የማህበረሰብ መከራና መጥፎ እድል የተጋረጠባቸውን በምን፣ እንዴትና መቸ ነገሮችን ለመቅረፍ የመታተር ስሜትና ምኞት ማንጸባረቅ መቻል ነው።
5	ፈቃደኛ (Volunteer)	ያሉትን ነገሮችን ለመስራት በፍቃደኛነት መነሳትና የህዝቡን ጥያቄና ሮሮ ሀሳብ ለመቀበልና መፍተሄ ለመስጠት በተሰፋ መሙላት መቻል ነው።

Table 2. 4 Offer Guideline

E.g. ***የውጭ ባለሃብቶች በተለይም ትውልደ ኢትዮጵያውያን ወደ አገሪቱ ገብተው መዋዕለ ንዋያቸውን እንዲያፈሱ የሚያስችል ሰፊ ጥረትና እገዛ እናደርጋለን*** => is the proposition (**P**) that the speaker can provide or propose an ideology for the audience as a negotiation. The audience desire the speaker intend to **P** as offering intention type.

V. Order

Order is as the speaker can give a command, enjoin, decree, rule, mandate and determine for once state information. On the other hand, the speaker can also give an instruct, direct, dictate, prompt and make provision to the transmitted information and also give an alert, declare, supervise, admonish, coordinate, govern, manage and be responsible for once action performed. In general, the user shall be belief to the speaker to take particular actions about the transmitted information.

No	The word infers to be	As the utterance can be reflected
1	ትእዛዝ (Order or Command)	ነገሮች እንዲስተካከሉ፣ እንስዲሰሩና እንዲፈጸሙ ጥያቄም ላቀረቡ ምላሻ እንዲያገኙና በተቻለ መጠን ለማቅረብ የሚተላለፍ መልእክት ማለት ነው።

2	ደንብ (Rule or Mandate)	ለማህበረሰቦች የሚተላለፍ ተቀባይነት ያለውን መረህና መመሪያ መቸ፣ እንዴትና በነማን መደረኛ መሆን እንዳለበት የሚገልጽና ምን እንደተፈቀደና እንዳልተፈቀደ የመግለጽ አዝማሚ መኖር።
3	ማስጠንቀቂያ (Alert, Admonish or Warn)	አሁንና ወደፊት ለሚኖሩ ችግሮች እንዲገነዘቡ፣ በወጡ መመሪያወችን እንዲያከብሩ የሚደረግ መልእክት ነው።
4	አስተምር (Instruction or Inform)	ስለሚሰሩት ነገሮች እውቀትን የማካፈልና አዳዳሲ ነገሮችን ለማሰልጠንና የተለያዩ ነገሮችን እንዲማሩ የማድረግ ሁኔታ የመፍጠር ነው።

Table 2. 5 Order Guideline

E.g. ***በአንዳንድ አካባቢዎች ከፍተኛ ቁጥጥር እንዲደረግና በማንኛውም ሕገወጥ እንቅስቃሴዎች ላይ እርምጃ እንደምንውስድ እናሳውቃለን*** => is the proposition (**P**) that the speaker can transmit an order or instruction or command for some groups to take action. The audience belief to the speaker to take a particular action **P** as ordering intention type.

VI. Advice

Advice is as the speaker can be give counsel, recommend, guidance, hints, tips, directions and offer suggestions about the transmitted information by other group members or politician. On the other hand, the speaker can also illuminate, commend, urge, promote, give notice, lead the way and give information about the transmitted message and also assist, steer, regulate, approve and big something up about the transmitted information. In general, the user shall be belief that the speaker give a guideline or recommendation for the transmitted propositions.

No	The word infers to be	As the utterance can be reflected
1	ምክር (Advice, Counsel or Recommend)	የውሳኔና የስነምግባር አካሄድን አስመልክቶ ምን ማድረግና ምን አለማድረግ እንደለለበት የሚሰጥ የምክር አገልግሎት ማለት ነው።
2	አስተያየት (Suggestion or Opinion)	በሰሩት ስራወች ላይ ምን እንደሰሩ፣ ምን መጨመራ፣ ምን መተው፣ ምን ማድረግ እንዳለባቸው የሚሰጥ ሀሳብ ነው።
3	ፍንጮችን (Give Hints or Tips)	ስለነገሮች፣ ስለተከሰቱ ክስተቶች ለማህበረሰቡ ወይም ለተፈለገው አካል ቀጥተኛ ወይም ቀጥተኛ ባልሆነ መንገድ ጠቃሚ መረጃወችን የመስጠት አዝማሚ ማለት ነው።

4	ማስተዋል (Heed or Espousal)	ስለአንድ ስራ፣ ሁኔታ፣ ጉዳይ እንዴት እንደሚሰሩና ሌሎች ቡድን እንዴት እንደሚሰሩት የምናጸባርቀበት የጥበብ ገጽታ ማለት ነው።

Table 2. 6 Advice Guideline

E.g. ***ለጉዳዩ ከባድ ትኩረት በመስጠት ፈጣን ርምጃ ልንወስድ ይገባል እንጂ የአንዳንድ አለመረጋጋት ለመላው አደጋ ነው*** => is the proposition (**P**) that the speaker can give a counselling or tips to the audience. The audience belief the speaker give a guideline or recommendation for **P** as an advice intention type.

VII. Request

Request is as the speaker can be solicit, appeal for, requisition, and implore the information to the other group members or politician. On the other hand, the speaker can also seek, importune, adjure, look for, endeavor, aspire, invite, summon, and also apply to other groups come to the transmitted information. In general, the speaker shall be give the user intends to the transmit propositions.

No	The word infers to be	As the utterance can be reflected
1	መጠየቅ (Solicit, Requisition or Ask)	ነገሮች እንዲሰሩልን፣ እንዲፈጸሙልንና እንዲታረሙ ለማድረግ ለሚመለከታቸው አካላት ጥያቄወችን የማቅረብና መልሶችን የመጠበቅ አይነት አመለካከት የሆነ።
2	ይግባኝ (Appeal or petition)	ለማህበረሰቡ ወይም ለራስ የሚሆን የገንዘብ፣ የመረጃ ወይም የተለያዩ የእርዳታ ነገሮችን ለሚመለከተው አካል የማሳወቅ ዘይቤ አቀራረብ ነው።
3	ፍላጎት (Demand or Require)	ለማህበረሰቡ ይሁን ለራሳቸው ስለአንድ ነገር (ጉዳይ) ትኩረት የመስጠት፣ አብሮ የመስራት ተነሳሽነትና ነገሮችን ለማከናወን የሚደረግ የበለጠ የመፈለግ ስሜትን የሚጭር ሀሳብ ማስተላለፍ መቻል ነው።
4	ጥሪ (Call, Beseech, Implore or Supplicate)	ለቆሙለት ማህበረሰብ፣ ለአባላቶችና ለመሳሰሉት የቅስቀሳ፣ የምክክርና የመሳሰሉትን ሀሳብ የማስተላለፍ ችሎታ መኖር ማለት ነው።

Table 2. 7 Request Guideline

E.g. ***አመራሮቹና አባላቱ በአስቸኳይ እንዲፈቱና እውነተኛ ወንጀል ፈጻሚዎች ተጣርተው ለሕግ እንዲቀርቡ እንጠይቃለን*** => is the proposition (**P**) that the speaker can solicit or petition to the respective

groups. The audience belief that the speaker can make the user intend to **P** as request intention type.

2.4. Sentence feature extraction

2.4.1. Sentence cleaning and preprocessing

In the spoken utterance there is not found any script character, numbers, and punctuation mark have been created from the speech. As we know, we have been processed textual data. We have removed any script character, numbers, and punctuation mark. Tokenization is the process of dividing the entire text into a word or sub-word or tokens. Tokenization of a sentence has a significant impact on the performance of text classification. Sentences are usually segmented with words or sub words by a morphological analyzer or byte pair encoding and then encoded with the word (or sub word) representations for the neural network (Tatsuya, Hiroyuki, & et al, 2019).

As stated earlier, we have used a deep learning approaches for invoking a huge amount of data for achieving better results in the proposed model. In the text classification procedure as the data set scarcity occurs to train the model. To maximum the data sets for text classification, there is applied a data augmentation techniques for fabricating other utterance depends on the original utterance. As general circumstance two data augmentation techniques are applied for natural language trained data. The one is easy data augmentation techniques that use simple heuristics to augment training data. The other one is back translation use noise introduced by neural machine translation to augment training data. In the easy data augmentation for fabricating other utterance, four techniques are there like synonym replacement, random insertion, random deletion, and random swap technique (Jason & Kai, 2019).

The synonym replacement is work for augmenting other utterance from the original once. It has works randomly choose n words from the utterance replace with synonyms. In the corpus, this N number of words can exist as vector values produce from the word embedding technique. To estimating the synonym of one word exists in the corpus, they can apply a cosine similarity between the selected words with other word exist in the corpus. They can apply both of the synonym replacement techniques like POS tag and threshold-based replacement mechanism (Konstantin , 2018). The random insertion is work for augmenting other utterance for randomly

insert synonym words in the utterance without the specification of words to be replaced in the utterance. The random swap is also working interchanging the number of words in the utterance as N time for fabricating other utterance format. It is simply exchanging the grammatical format of the utterance. The random deletion is performed remove words from the utterance up to the specified probability of to give remove from the utterance.

2.4.2. Word Embedding

We have to represent the words in a numeric format to be understandable by the computers. Word embedding is one of the most dominant in deep learning approaches as feature extraction from textual data sets. The word embedding can be trained using the neural network language model. There are different word embedding techniques for representing the raw text into a numeric (vector) format like word2vec, glove, fast text, and contextual word representations (François, 2018).

2.4.2.1. *Word2vec*

Word2vec is one of the word embedding techniques for converting the text into its corresponding vector-matrix representation, it is first proposed by (Mikolov, Kai , & et al, 2013)**.** Word2vec reconstructs the linguistic context of words, these means the word or sentence surrounding spoken or written language (disclosure) helps in determining the meaning of the context, and also the word2vec learns vector representation of words through the contexts (François, 2018)**.** Word2vec is based on neural networks which were a new approach for language models in the time of this technology was published. There is such technique like N-grams, which build matrices overall words or a window of n words, was heavily used before but relied on a lot of high-quality data which isn't always at reach for analysts. Therefore, new technologies requiring less data with the same performance had to be built.

In word2vec there are two architectures for representing text into dense low dimensional matrix representation and learned from the data. The two architectures are a continuous bag of a word (CBOW) and a skip-gram model. In both architectures, there is exist three layers the input layer that accepts the input value, hidden layer for projection view, and output layer for predicate value (vector representation) of the word. The word representations are a critical component for many natural language processing systems to represent words as indices in vocabulary. This

word representation cannot capture the rich relational structure of the lexicon words that exist in the vocabulary. The vector-based models are provide a better extraction of word lexicon in the vocabulary. They encode continuous similarities between words as distance or angle between word vectors in a high-dimensional space (Andrew , Raymond, & et al, 2011)

The CBOW model can predict the current word from the surrounding context words. It's mainly applicable for predicting missing words in the sentence, give meaningful n-gram information, and also effective for sentiment orientation. The CBOW has taken the context words to predicate the target word, as the input layer has taken the context words it depends on the size of the window of the target words corresponding to left and right directions. In the hidden layer shows the projection view of the words in the vectors. Finally at the output layer can yield the target word values in the corpus vector representation. It has taken as its inputs the high dimensional one-hot vectors of words in the corpus and produces a vector space of several hundred dimensions which are much smaller than the size of the vocabulary, such that each unique word in the corpus is represented by a continuous dense vector in the embedded vector space (David M. , 2016).

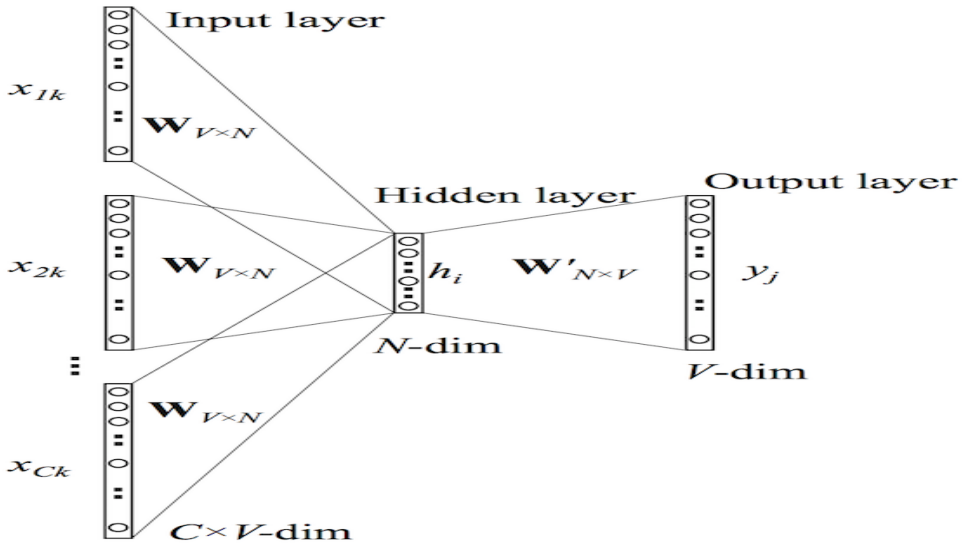

Figure 2. 1 CBOW architecture

The skip-gram model has been predicting the corresponding context words from the given words. It's mainly applicable for knowing the nearest word in the sequence, semantically or

logically related words. The skip-gram has taken as input target word and yield the context word at the output layer. As the input layer takes the target words as one-hot encoding structure, at the hidden layer they can show the projection vector of the word, and finally in the output layer can yield the context words depend on the window size to its target word. They said that, the model trains word embedding by maximizing the probabilities of context words given their context windows for finding the neighboring words in the vocabulary (Peng, Yue, Xingyuan, & Yunqing, 2016).

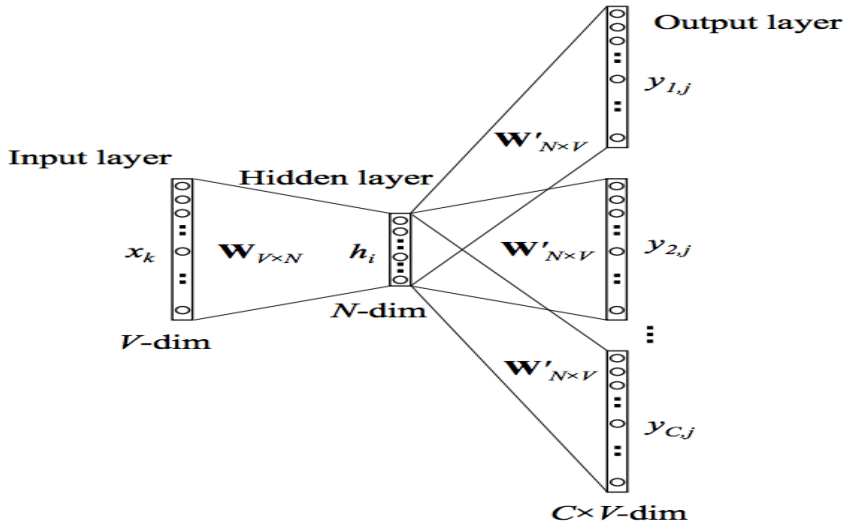

Figure 2. 2 Skip gram architecture

2.4.2.2. GloVe

Recently, many methods to generate word vectors are based on statistics regarding their surrounding context words and are separately trained on them. The GloVe is an unsupervised learning algorithm for obtaining vector representation for words. In Glove training is performed on aggregated global word-word co-occurrence statics from a corpus and the resulting representation showcase interesting linear substructures of the word vector space. First, it has established some notation. Let the matrix of word-word co-occurrence counts be denoted by X, whose entries X_{ij} tabulate the number of times word j occurs in the context of the word i (Jeffrey, Richard, & Christopher, 2014).

Let $Xi = \sum_{k=1}^{n} X_{ik}$ be the number of times any word appear in the context of the word i. for finding the probability that word j appears in the context of word i. $P_{ij} = \frac{X_{ij.}}{X_i}$ Or in other words,

Glove minimizes the objective function $J = \sum_{ij=1}^{v} f(X_{ij}) * (w_i TW_j + b_i + b_j - \log(X_{ij}) ** 2$ to find its word vectors. Finally, the weighting function expressed as follow as:

$$f(X_{ij}) = \begin{cases} \left(\frac{X}{X_{max}}\right) ** a \\ 1 \end{cases} \quad if\ X < X_{max} \quad otherwise \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

Finally, the author's would be concludes that Glove, is a new global log-bilinear regression model for the unsupervised learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks.

2.4.2.3. FastText

The FastText is a library created by the Facebook Research Team for efficient learning of word representations and sentence classification. It is helpful to find the vector representation for rare words and also give the vector representations for the words not present out of vocabulary words in the dictionary. It has better for computing small datasets as the performance of semantic and syntactic for words in the sentence. According to (Piotr, Edouard, Armand, & Mikolov, 2017), the FastText can take the morphology of words into account, each of the word consists of multiple characters and the combination of character. It represents a word by its ID within a vocabulary and its n-grams of characters. The whole representation of a given word is the sum of all vectors of the n-grams contained in a word.

$$W_i = \sum_{g \in Gi}^{i} Z_g \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

The W_i being the ID of a word, G is the set of n-grams and Gi the set of n-grams of word i and Zg the specific vector for n-gram g.

The FastText is better for textual data classification as a deep neural network model trainer for much high representation of words in the corpus. As they said, the FastText can decrease the prerequisite preprocessing stage for text classification and also better representation of word values (Birol, Cuneyt, & Selman, 2019).

2.4.3. Sentence embedding

As word embedding has to embed word in the vector representation, they can also represent sentence as vector representation for finding the semantic relationship between sentences and also take the sentence values for text classification purpose like for analysis the comment of sentiment, news classification, and also document classification (Mridul & Jaydeep, 2019). In sentence embedding, there are different techniques for embedding each sentence into vector format like mean embedding, Doc2vec, InferSent, skip thought, and universal sentence encoder techniques.

2.4.3.1. Mean embedding

The mean embedding is a supervised learning method that depends on the word embed vectors for word representations of each word and taking the average value of each word vector value. In order to generate the sentence vector value first, there is an available vector value for each word generated by the word embedding techniques of each word in the corpus. Finally, it has taken the average vector weight value of words in the sentence for representing the sentence vector values (Jey & Timothy, 2016).

$$Se = \sum_{j=1}^{n} \frac{W_j}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (2.3)$$

This is for finding sentence vector (Se), n is number of words exist in the sentence, Wj is the vector values of each word.

2.4.3.2. Skip-Thought Vector

The Skip-Thought Vector is an unsupervised learning method for sentence embedding. This technique can predict the neighboring sentence or phrase for the given sentence. The analogous framework of skip thought is the same as the word2vec skip-gram models. The techniques can be applied in recurrent neural networks for classification purposes (David, Ben , Wei , & et al, 2006). According to (Ryan, Yukun, & et al, 2015), they can treat skip-thoughts in the framework of encoder-decoder models. The encoder has used to map words to a sentence vector and also the decoder has used to generate the surrounding sentences. The encoder-decoder model has gained a lot of traction for neural machine translation. In this setting, an encoder is used to

map e.g. an English sentence into a vector. The decoder then conditions on this vector to generate a translation for the source English sentence.

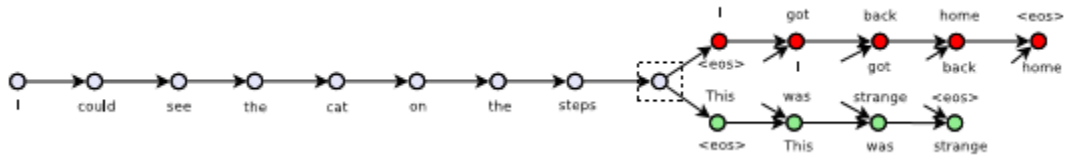

Figure 2. 3 Skip-thought vector

2.4.3.3. Doc2vec

The Doc2vec is an unsupervised framework that learns continuous distributed vector representations for a piece of text. The texts have been a variable-length, ranging from sentences to documents. This technique has been inspired by the learning vector representations of words using neural networks that the word vectors are asked to contribute to a prediction task about the next word in the sentence (Mikolov & Quoc, 2014). From this idea, this method also asked to contribute to the prediction task of the next word given many contexts sampled from the paragraph.

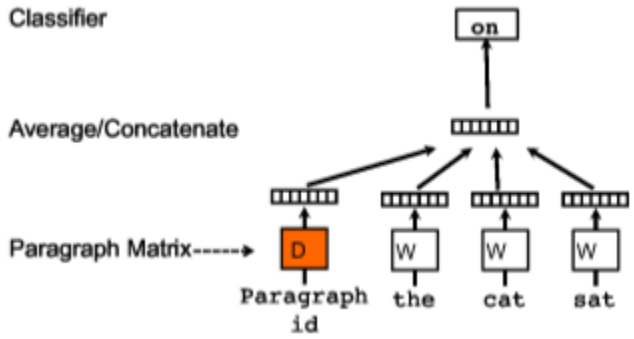

Figure 2. 4 Doc2vec

In the above figure shows, every paragraph has mapped to a unique vector, represented by a column in matrix D and every word is also mapped to a unique vector, represented by a column in matrix W. The paragraph vector and word vectors are averaged or concatenated to predict the next word in a context. In the experiments, they have used a concatenation as the method to combine the vector value (Mikolov & Quoc, 2014).

2.5. Convolutional neural network

Deep learning is about deeper neural networks that provide a hierarchical representation of the data by means of various convolutions. This allows larger learning capabilities and thus higher performance and precision values to be achieved (Alom, Tarek, & al, 2018). We have selected one of the most dominant classifications deep learning approach is CNN in NLP for text, sentence, sentiment, and document classification tasks. The CNN approach has been first proposed by (Fukushima, 1988)**.** This approach has some characteristics as good result in various field like the weight sharing of network structure are more similar to the biological neural networks, reduce the complexity of the network model, a reduction in the number of weights and the direct use of any objects to avoid hand-crafted feature.

The CNN architectures are composed of two-phase like feature extraction and classification phase. In the feature extraction phase, there is a four-layer like a convolutional, rectified linear unit (ReLU), pooling, and fully connected layers. In the classification phase, there are trained the model and classifier phase by using the different activation functions (John, 2016). In the CNN approach mainly two challenges can happen for predicting the training and testing accuracy after training the models. Under fitting occurs when the training error can be high as the testing error, this conditions can occur the number of input data is less for training the models, for compensating this errors they have used data augmentation techniques for enlarging the training data. Over fitting occurs when the testing error can be high as training error, this conditions can happen the number of input data is more for training the models, for this time they have applied a dropout or regularization techniques to remove some value of weight at training time and also adding the remove weight value in the testing phase (Gavrilov, Jordache, Vasdani, & Deng, 2018).

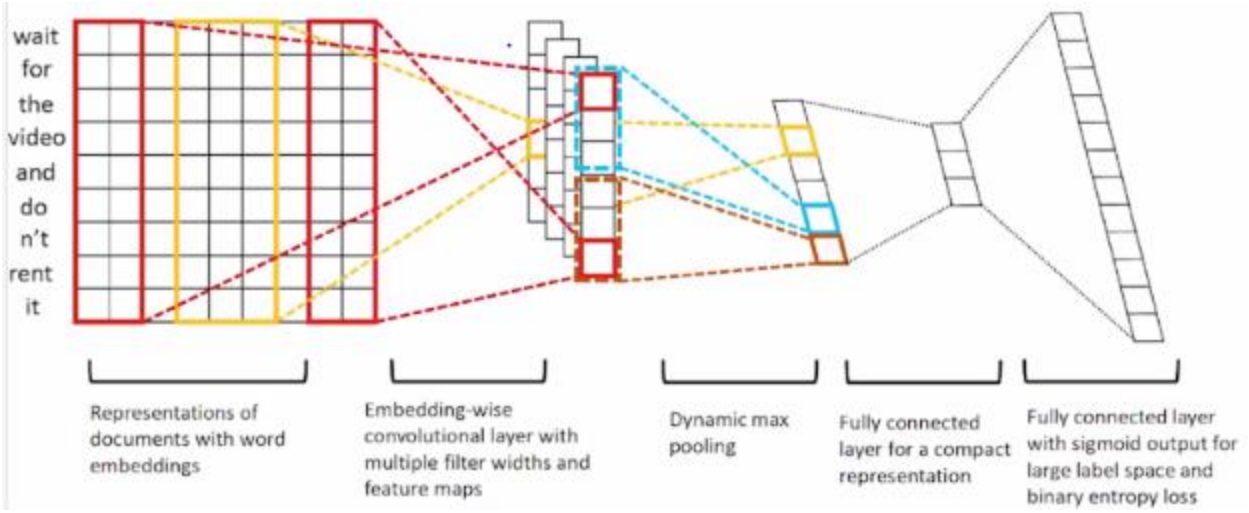

Figure 2. 5 CNN architecture

2.5.1. Component of CNN

The component of CNN have been two phase the first phase is feature extraction. In the feature extraction mainly used four layers to converting the matrix representation to one array value like convolutional, rectified linear unit, pooling and full connected layer. The second phase is classification phase, in this phase they can take one dimensional array values for training the network and also applying different activation function for estimating the output values.

2.5.1.1. Feature Extraction

One of the most interesting classification algorithm from other machine learning or deep learning approach. The CNN are deeply extracting pattern features from the input objects as performing a numerical values with huge amount of data sets (John, 2016). We have discussed the four commonly layer performed on the CNN architecture.

A. Convolutional layer

In the convolutional layer, we have taken the matrix representation of sentence as input and also filter (kernel) size with matrix size of 3x3 or 5x5 for gaining a feature map for the matrix representations of the input values. The steps perform in convolutional layers are, first multiplying the corresponding matrix representation by the filter size, next we can sum up each value (Prabhu, 2018).

As an example take matrix representation of size 7x7 as input to the first layers with filter size of 3x3 for gaining feature maps from the inputs. In the convolutional layer there is padding and stride. Padding is increase the matrix boarder by adding 0 in all positions and stride is the movement of filter size after performing operation in the matrix representation like from vertical or horizontal. In this example there is not consider padding value and also not given the stride values by default it can takes 1 for both padding and stride.

1	2	0	1	2	1	0
1	2	0	3	1	0	1
0	3	1	2	1	1	1
1	2	0	3	1	0	1
0	3	1	2	1	1	1
1	2	0	1	2	1	0
1	2	0	3	1	0	1

1	0	1
0	1	0
1	0	1

Figure 2. 6 Dataset and filter size

On the above shows left side the input matrix representation of the sentence with size 7x7 and on the right side there is filter (kernel) size values with size 3x3.

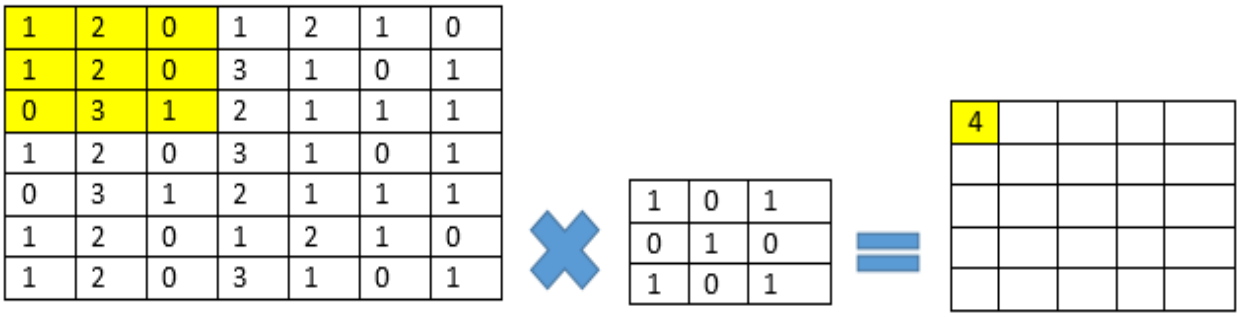

Figure 2. 7 First convolution

On the above shows that the input matrix representation can multipy by the corresponding filter size, next sum up the values we give feature map value 6.

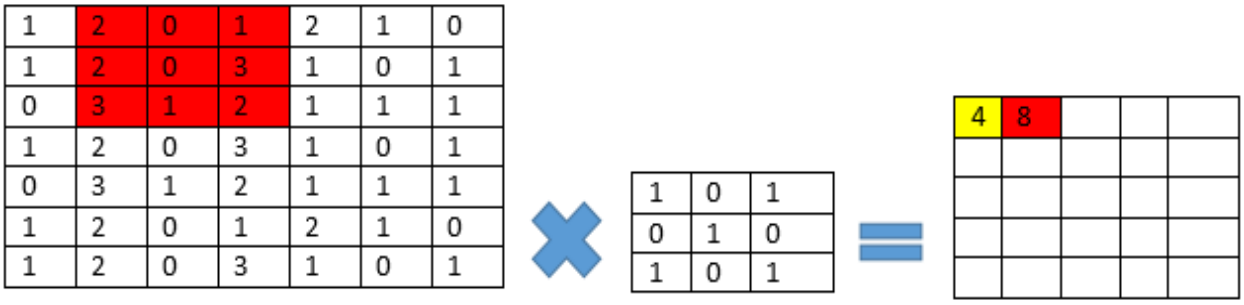

Figure 2. 8 Second convolution

On the above shows that the input matrix representation can be multiple by the corresponding filter size by moving stride value 1 to horizontal poition, it will give feature map value 10.

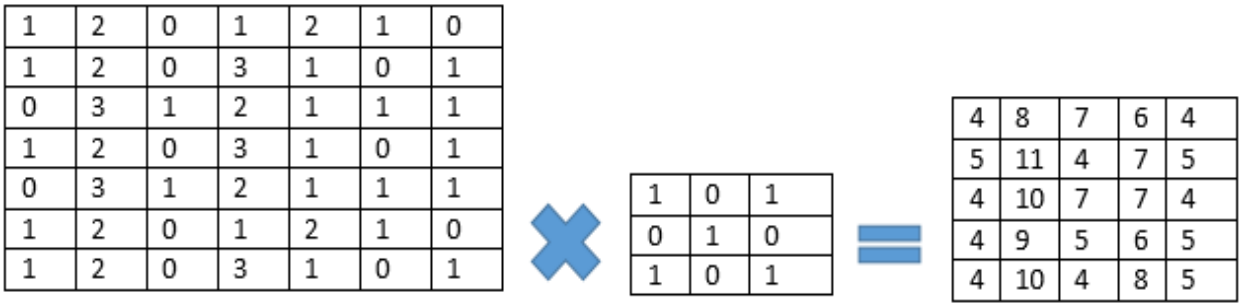

Figure 2. 9 Final result

We have continue above step for computing each feature map from the matrix representation finally, we have gain the above right side feature map values from the left side input.

In general, let the input matrix representation is Si, the filter size is Fi and the output feature map value is X.

$$X = \sum Si * Fi \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.4)$$

B. Rectified linear unit (ReLu)

The ReLu function have remove every negative values from the filtered feature map values by replacing this by zero. To evoke the ReLu function to avoid the values come to be zero (Prabhu, 2018).

$$F(x) = \begin{Bmatrix} 0 \; if \; x < 0 \\ x \; if \; x \geq 0 \end{Bmatrix} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.5)$$

C. Pooling layer

In the pooling layer, after performing the ReLu function on the feature maps next step is pooling the feature map values. The pooling layer has also reduced the dimensionality of input features from the feature maps to enable reduce the number of the input parameter and training computational time (Dshahid, 2020). In pooling layer commonly uses two pooling mechanisms like max and average pooling. They have applied the pooling layer, first specify the window size that wants to access from the feature map, the stride after performing pooling the horizontal and vertical movement in the feature map, next it can go the window across the feature map, and finally from each window, it can take the max or average values as they wanted.

In the case of the average pooling, this function usually sums up over N×N patches of the feature maps from the previous layer and selects the average value. On the other hand, in the case of max-pooling, the highest value is selected from the N×N patches of the feature maps. Therefore, the output map dimensions are reduced by n times (Alom, Tarek, & al, 2018).

As an example we can take a 5x5 matrix value with a window size 2 and stride value 2. On the right side shows the max pooling values from the feature maps and in the bottom shows the average values from the feature map.

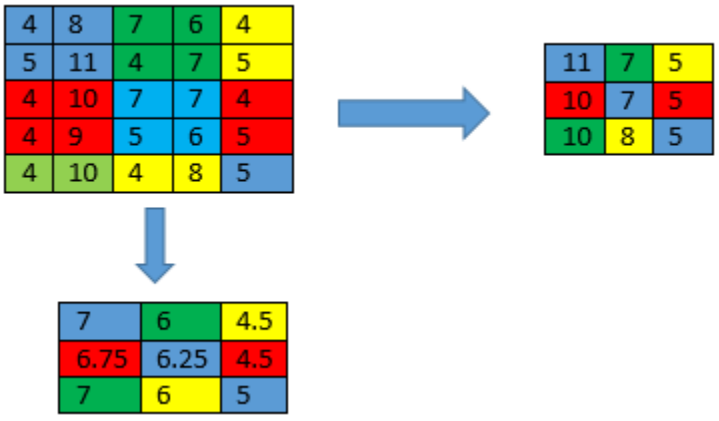

Figure 2. 10 Pooling layer

D. Fully connected layer

The fully connected layer can be accept the outputs of the pooling layer puts them in the single vectors. This value like flatten one dimensional array list like value to feed the next neural network.

2.5.1.2. Classification phase

In the classification phase of CNN architecture after gaining the fully connected value, it has followed a feed forward neural network like structures for training the model with input to weight values. It has performed a multiplication of input value with weight value, after that it has applied a summation function and also finally applied a learning function for estimating the output value of the training data. There is exists a number of dense layers for optimizing its value to training the network in wise use manner. Mainly it can used a non-linear functions to allow the model create complex mapping of the network input and output layer, for learning and modeling a complex data, it can also allow back propagation in the network to learn and reduce the error, stacking in different neuron and to normalize the output range (Rikiya, Mizuho, Richard, & Kaori, 2018). Some of the activation function performed in the CNN approach.

I. The *Rectified linear unit (ReLu)*

The ReLU (rectified linear units) function, this is computationally efficient, it looks like linear functions as derivative function and also allows a back propagation. As the value approaches to zero the network not learn a back propagation. Much faster to learn, few network have been activated. The estimated values range are 0 to positive infinity (Jason B. , 2020).

They can improve the performance rate of CNN classification on by using ReLu inside the fully connected layer of CNN for the experiment result of Javanese vowels sound (Chandra & Afiahayati, 2018). They can also use ReLu as classification in the deep learning by estimating the threshold values in the output values (Abien , 2019).

In the training phase of the network model the ReLu activation function perform in the following equation (2.7).

$$f(x) = \max(0, x) = \begin{Bmatrix} xi, \ if \ xi \geq 0 \\ 0, if \ xi < 0 \end{Bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

After trained the model, for update the weight and bias value by calculate the error values in the specified layer they can require the derivative value of ReLu.

$$f(x)' = \begin{cases} 1, if\ xi \geq 0 \\ 0, if\ xi < 0 \end{cases} \quad \text{......} \quad (2.8)$$

II. Sigmoid

The Sigmoid (logistic) function, this is a smooth gradient, the estimated values between 0 and 1, computationally expensive, vanishing gradient problems happen. It mainly uses in the output layer for classification purposes. They have been used the sigmoid functions in the feed-forward neural networks for reducing the error signal problem, oscillation problem, and the symmetrical problems that occurred during training the models and in the back propagation stage for updating the weight value. In the sigmoid function invoking in the hidden layer, the random initialization weight values can be minimum for reducing the vanishing gradient problems (Xavier & Yoshua, 2010). In the neural network design component, the hidden layer and output layer uses a transfer function for predictive values. They can apply the sigmoid nonlinear activation function in the hidden layer for reducing the error rate, to learn any complex feature data, handling noise data, and improving the predictive accuracy (Dhana , 2019).

In equation 2.9 express the sigmoid function values for estimating in value in the output layer.

$$f(x) = \frac{1}{(1+ e^{-x})} \quad \text{......} \quad (2.9)$$

After trained the model, for update the weight and bias value by calculate the error values in the specified layer they can require the derivative value of sigmoid functions.

$$f(x)' = \frac{1}{((\frac{1}{1+e^{-x}})(1-\left(\frac{1}{1+e^{-x}}\right)))} \quad \text{......} \quad (2.10)$$

III. *Softplus*

Softplus is another activation function mainly used in the hidden layer of neural networks. It's mainly preferable property is differentiable and its derivative is easy to demonstrate to each corresponding layers of the neural network. It is unbounded activation function that has been preserving the problems occur for other activation function in the hidden layer. The problems mostly occur like the weight gradient, which is used to update the weight in the back propagation trend, and also vanishing gradient problems as the model became deeper. They have applied softplus activation function for reducing vanishing gradient problem in the deep neural network,

they have achieved better phone error rate than sigmoid and ReLu in the recognition of phoneme tasks (Hao , Zhanlei , & et al, 2015).

As trained the network using softplus based activation function the output value from the hidden layer can be passed by the softplus function as input 'x' and output 'f(x)' its estimated value range can be between (0, positive infinity):

$$f(x) = \log(1 + e^{x}) \quad \text{..... (2.11)}$$

After trained the model, for update the weight and bias value by calculate the error values in the specified layer they can require the derivative value of softplus.

$$f(x)' = \frac{1}{1+ e^{-x}} \quad \text{..... (2.12)}$$

IV. *Softsign*

The softsign activation function can be applied in the hidden layer improving the performance accuracy for handling vanishing gradient problem (Fatih & Galip, 2017). They have used the softsign activation function for limits the range of the output value in a text to speech converter for avoiding the saturation problem that exponential based nonlinearities sometimes happen, they can also initialize the convolution filter weights with zero-mean and unit-variance activation throughout the entire network model structure (Wei, Kainan, & et al, 2018). For training the models by using a softsign activation functions, they can convergence to the polynomial values equation (2.13) express.

$$f(x) = \frac{x}{(|x|+1)} \text{..... (2.13)}$$

After trained the model, for update the weight and bias value by calculate the error values in the specified layer they can require the derivative value of softsign.

$$f(x)' = \frac{x}{((1+|x|)**2)} \text{..... (2.14)}$$

V. *Swish*

The swish is a new activation function with the properties of one-sided bounded at zero, smoothness, and non-monotonicity in the neural network. As performing in the deep neural network environment the ReLu function is the default for compensating vanishing gradient

problem form other activation functions. It is numerous ongoing illustration have been proceeding for replacing the ReLu function with other activation function, but the improvement of performance have been varied as using in different model and dataset. They have been proposed a swish activation function for compensating this stated problems as a solution in a rational way across in different model and dataset (Prajit, Barret , & Quoc, 2017).

In the training process of model, for applying a swish activation function we can express in equation (2.15) with input 'x' and output value 'f(x)'.

$$f(x) = x(1/(1 + e^{-x})) \quad \text{..... (2.15)}$$

After trained the model, for update the weight and bias value by calculate the error values in the specified layer they can require the derivative value of swish in equation (2.16), where **x** is input value and **y** is output values from the network.

$$f(x)' = y + (1/(1 + e^{-x})(1 - y) \quad \text{..... (2.16)}$$

VI. *Softmax*

The SoftMax function is a type of sigmoid, mainly able to handle multiple class values, useful for in the output layer of neurons. The Softmax function produces an output range of values between 0 and 1, with the sum of the probabilities been equal to 1. This function mostly appears in almost all the output layers of the deep learning architecture. The softmax activation function can be applied in different sectors for multi-class values and the data sets are more mutually exclusive for differentiating one from the other (Radhika , Bindu , & Latha , 2018). The softmax can be better classifier for CNN from the other integrate machine-learning algorithm to classification purpose, as providing predictive probability for each speech act classes (Yoo, Ko, & Seo, 2017).

In the training process of model, for applying a softmax activation function we can express in equation (2.17) with input 'x' and output value 'f(x)'.

$$f(x) = \frac{e^x}{\sum_{i=0}^{n} e^{xi}} \quad \text{..... (2.17)}$$

After trained the model, for update the weight and bias value by calculate the error values in the output layer they can require the derivative value of softmax in equation (2.18).

$f(x)' = (e^{xi}\ (\sum_{k=1}^{n} e^{xk}) - (e^{aj}\ *\ e^{ai}))/(\sum_{k=1}^{n} e^{ak})\ **\ 2 \ldots\ldots\ldots\ldots\ldots\ldots$ *(2.18)*

2.6. Performance metrics

In this study, for evaluating the system performance we have used some metrics to knowing the trustful of expected and predicted values for invoking the data set with the learning models. We have used accuracy metrics for evaluating training accuracy by giving the training data and also testing accuracy that is the percentage of test set samples that are correctly classified by the model learned by the training samples. We have also used a confusion matrix, it is a useful tool for analyzing how well your classifier can recognize tuples of different classes of the training model. The confusion matrix displays in graphical format the number of correct and incorrect predictions label values made by the proposed model compared with the actual label in the test dataset.

	Predicted class	
Actual class	C_1	C_2
C_1	True positive(TP)	False negative(FN)
C_2	False positive(FP)	True negative(TN)

Figure 2. 11 Confusion Matrix

True positive (TP) is the number of true positive. It is the positive tuples that were correctly labeled by the model classifier. The class label is assertive correctly predicted as assertive

True Negative (TN) is the number of true negative it is the negative tuples that were correctly labeled by the classifier.

False Positive (TP) is the number of false positive tuples. It is a negative tuples that were incorrectly labeled by the classifier a positive tuples. As sample label assertive as predicated as other labels

False Negative (FN) is the number of false negative tuples. It is a positive tuples that were incorrectly labeled by the classifier a negative tuples. As sample with directive as predicated as other labels.

Accuracy is reflecting the reality of the model to be achievable in measurement value by considering a bias. It is reflecting a measurement values that the model predicators are close to the truth values. In equation (2.19) reflects the accuracy measurement calculation.

$$A(accuracy) = \frac{Number\ of\ correct\ classification}{Number\ of\ test\ cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (2.19)$$

We have also used a classification report metric of recall and precision for knowing the accuracy performance of each class. The precision is simple what percentage of tuples labeled as positive as actual as such for measuring exactness of the model. The general equation for expressing precision values in equation (2.20). Its target is minimizing the false positive values.

$$Precision = \frac{TP}{TP+FP} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (2.20)$$

The recall is simple what percentage of tuples labeled as positive as labeled as such for measuring the completeness of the model. Its target is minimizing the false negative values. In equation (2.21) can be express the recall.

$$Recall = \frac{TP}{TP+FN} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (2.21)$$

We have also used the F – measure for knowing the performance of the model. The F – measure more favorable for expressing the model measurement values. It can be combine both of the recall and precision values from the model. The equation (2.22) can be express the F – measure value as in general.

$$F\ measure = \frac{2*(precision*recall)}{precision+recall} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (2.22)$$

2.7. Related works

There are some related works in different perspective areas concern for speech act and intent recognitions, by applying rule-based, machine learning, deep learning approaches from textual, speech, and multimodal data. As the main concern with the textual data for recognizing action-based speech act and intent from a human speech utterance. According to (Nikolay , Alexander, & Alexey , 2017)**,** they have developed a model to predict intention using deep learning approaches, for the analysis of an illocutionary act in political discourse from social media for the classification of intention type and direction of the intention. They have applied a word

embedding techniques for training the models for both the neural networks with a huge amount of textual data, finally, they have suggested as the number of class increase long short term memory are more preferable for text classification than CNN based on taking some classification metrics on the experiment. However, the transfer function is important for estimating the output value and predicating its accuracy for deep learning approach. In this work not clearly defined which transfer functions can be applied for predicting the intention and direction of intention accuracy with these number of class.

According to (Shivashankar, Trevor, & Timothy , 2019), they have developed a target based speech act classification in political campaign text by segmentation of each sentence with respective speech act and target party. They have used an Elmo embedding with the model of BiGRU forward and backward techniques for represented a vector representation of words, with a word dropout semi-supervised approaches for classification of speech act and corresponding target party. They have compared the proposed approaches with others by using some classification metrics as evaluations. The authors have predicated the speech act towards the target party (to whom we have spoken). But, for the target party, they can create a misconception ideology from the transmitted message (the target party did not find the goal or intent of the speaker).

According to (Kyoungman & Youngjoong, 2018)**,** they have developed a speech act classification by applying a word embedding CBOW and ranking based word embedding (RBWE) features extraction techniques for encoding the lexical features of unigram and discourse features from the utterances, finally, they have achieved the state of the art results by applying a CNN classifier with word embedding techniques for speech act classifications. But, they cannot reflect once the speech act class to its branch for the intended users as create well-defined interaction as in the hierarchical format and also which transfer function performs for justifying the word2vec model.

According to (James , Zuhair, & Keeley , 2010), they have been classified speech act into question or non-question, as feature extraction of utterance replacing each content word with a standard numeric wildcard token and the salient features word by replacing function words with numeric tokens as a classifier of the decision tree has been achieved better results. however, as the number of class enlarge these feature extraction technique to this classifier not converge to

optimal results, because of this, we used deep embedding techniques for extracting unique features to the utterance as an enlarging data set for covering its classification accuracy reliable we can use a deep learning approach.

According to (Junmei & William, 2019), they have analyzed a phone call transcripts for predicting customers call intent from the caller transcripts data with a scalable data labeling methods for labels the four class of each sentence. They have represents each sentence with semantic features and into matrix-vector representation uses the pre-trained word embedding techniques of both word2vec and GloVe as fusion fashions. They have developed an AI-based customer caller transcript as training with CNN for multi-class classification with a quantitative metrics for evaluation criteria.

CHAPTER THREE

3. SYSTEM DESIGN

3.1. Introduction

In this chapter mainly, we have to be considered and discussed the overall system model of speech act and intent classification tasks for Amharic political speech. In section 3.2, we have view the general system model overview and also design the system models for classifying the Amharic speech act and intent. Next, we have driven and also give a detailed explanation about each sub-modules of the system design flow chart like in subsection 3.2.1, we have discussed the data preprocessing for cleaning the raw data set and data augmentation techniques for fabricating other utterance. In subsection 3.2.2, we have discussed the detailed representation of each word value convert to vector format by using a word embedding approach with the selected technique.

In subsection 3.2.3, we have discussed a sentence embedding technique for representing each utterance to a vector format. Finally, in subsection 3.2.4, we have discussed the CNN architectures for both feature extraction from the vector value of each utterance and also analyzed the activation functions in the classification phase.

3.2. Proposed speech act and intent classification model

The proposed system model has yielded the manual system flow charts as a mechanism for problem-solving. The systems have been performed a speech act and intent classification for Amharic political texts. Diagram 3.1 shows the general functionality of the Amharic political speech act and intent classification framework. The models have been four main individual components.

In the first component, we have performed a data preprocessing stages. There is an Amharic text corpus for describes and state in the political situation for preparing and will suited conditions we have applied a data cleaning to remove unwanted script, punctuation mark, and numbers and also stop words. We have used data augmentation techniques for fabricating other utterance to overcome the scarcity of training data.

In the second component, we have selected and discussed the word2vec technique of CBOW architecture for converting each word into the corresponding matrix representation value. In the third component, we have selected and discussed a mean embedding technique for computing each word values of the utterance.

The fourth component is a deep learning CNN approach that has been applied for both feature extraction and classification phase of Amharic speech act and intent classification task. The below diagram represents each of the four components as one integrated view, and also discussed each component separately and how it performed.

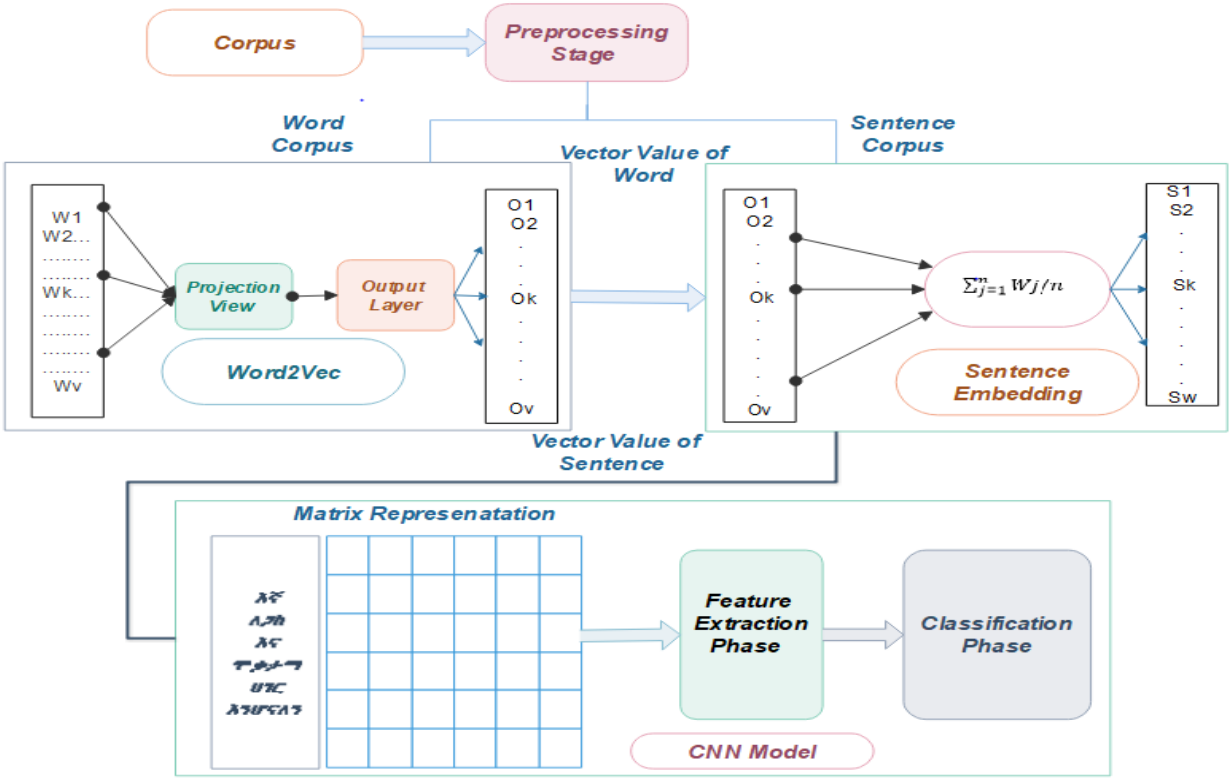

Figure 3. 1 Proposed system model

3.2.1. Data Preprocessing

As we have collected the Amharic political text from different politicians stated manifesto, political party debate from different studio programs, and also from social media as text format. The first step is to remove punctuation mark, numerical values and any style of words exist in the corpus, secondly, we have to remove silence or unwanted words from the corpus that will not contribute any percentage values for the stated political texts and also reformulate and ready for working purpose the corpus. Finally, we have tokenized the entire corpus text into a unique word for representing an encoding value in the entire corpus. This encoding values used for input values to the neural network in the input layer.

As the number of utterance data scarcity has happened as numbers of the class can be enlarged for coverage of these diverge values, we have selected synonym replacement data augmentation technique. First, we have to collect action verbs in the corpus that can give some responsibility for that utterance. Secondly, we have performed a cosine similarity of the selected word from the entire corpus for replacing the word by other words in the corpus. Finally, we have performed

tokenization of sentence into words for replacing the position action verb words by other words exist in the corpus for formulating another utterance.

For instance, we have taken one utterance from the corpus for augmenting other utterance depends on this utterance. "ሚድያዎቻችንና ጋዜጠኞቻችን እዉነትን ለማረጋገጥ አይሰሩም "It is the original utterance. In this utterance, the action verb is "አይሰሩም" the selected word from the utterance that has been replaced by another word from the corpus. This selected word "አይሰሩም" has been passed to the word2vec model to perform the cosine similarity between other word vector values in the corpus, after performed this operation we have selected most similar related word in the corpus-based on a vector value.

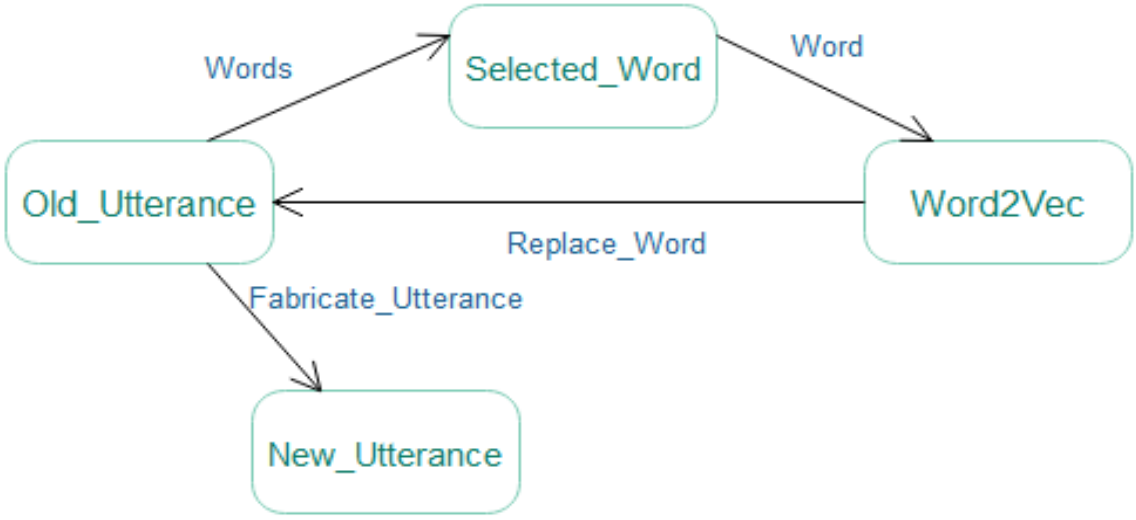

Figure 3. 2 Synonym replacement data augmentation technique

3.2.2. Word embedding

In this stage, we have selected and discussed a word2vec technique as a neural network language structural model for representing each word in the corpus with a vector value.

3.2.2.1. Word2vec

The word2vec is one of the word embedding techniques for converting each word to vector format by applying a neural network. In the word2vec two architecture are exist for performing this activities, we have selected the CBOW architecture for representing each word's value in the corpus. The CBOW could take the corresponding context words an encoding values as it depends on the window size for predicting the target word vector values (David M. , 2016). As,

the corpus contains the entire words like [W1, W2, Wk.… Wv] as W is the words exit in the corpus, Wk. is the target word in the corpus and V is the size of vocabulary for the corpus. As finding the target word value Wk., we have taken the consecutive left and right context word one hot encoding values as input values for the neural network.

Parameter specification for word2vec	
Window size	2
Learning rate	0.01
Embedding size	256
Architecture	CBOW
Min count of word	1

Table 3. 1 Parameter specification for word2vec model

As, a target word values Wk. = [0, 0, 1, 0, 0 ………..0]

As, a left context word value Wk-2. = [1, 0, 0, 0, 0… 0]

As, a left context word value Wk-1. = [0, 1, 0, 0, 0… 0]

As, a right context word value Wk+1. = [0, 0, 0, 1, 0 …0]

As, a right context word value Wk+1. = [0, 0, 0, 0, 1 …0]

As, N is the number of hidden layer exist in the neural network.

As, a weight one values between the input and hidden layer with size (V, N). W1 = (V, N)

As, a weight two values between the hidden and output layer with size (N, V). W2 = (N, V)

The hidden layer (H) value is the dot product of weight one transpose with input values (X). Equation (3.1.) shows the output values as hidden layers. The output layer (y) value is the dot product of weight two transpose with hidden layer value. Equation (3.2) finding the output values in the output layer.

$$H = W1.T * X \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (3.1)$$

$$Y = W2.T * H \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (3.2)$$

As, finding the target word values for the input word values, we can apply a SoftMax activation function as the output layer for predicating the values. The yi is the output values from the output layers. Equation (3.3) performs the SoftMax function for estimating its vector value.

$$S(yi) = \frac{e^{yi}}{\sum_{i=1}^{n} e^{yi}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.3)$$

As for can converge the target word values we can back propagate the error rate with update weights between the each layer of the neural network with the learning rate value. We have performed the derivatives of SoftMax for updating the initial weight value. In equation (3.4) express the derivative SoftMax output value with respect to the output layer values.

$$\frac{dSi}{dOj} = \frac{\left(\frac{e^{yi}}{\sum_{k=1}^{n} e^{yk}}\right)}{dOj} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.4)$$

In equation (3.5) shows the differentiable of SoftMax values, (3.6) shows that by applying in the forms of quotient formals derivatives applied, (3.7) shows that the derivatives of output layer values to its output value, and also (3.8) shows that the final change values by applying a SoftMax.

$$dSi = \left(\frac{e^{yi}}{\sum_{k=1}^{n} e^{yk}}\right) * \left(1 - \frac{e^{yk}}{\sum_{k=1}^{n} e^{yk}}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.5)$$

$$dSi = Si * (1 - Sk) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.6)$$

$$dOj = dij \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.7)$$

$$\frac{dSi}{dOj} = \frac{(Si * (1 - Sk))}{dOj} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.8)$$

After calculate the change value, we can calculate the weight change values for train the network models repeatedly.

$$W1 = W1 - (learning_{rate} * \left(\frac{dSi}{dOj}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.9)$$

$$W2 = W2 - (learning_{rate} * \left(\frac{dSi}{dOj}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.10)$$

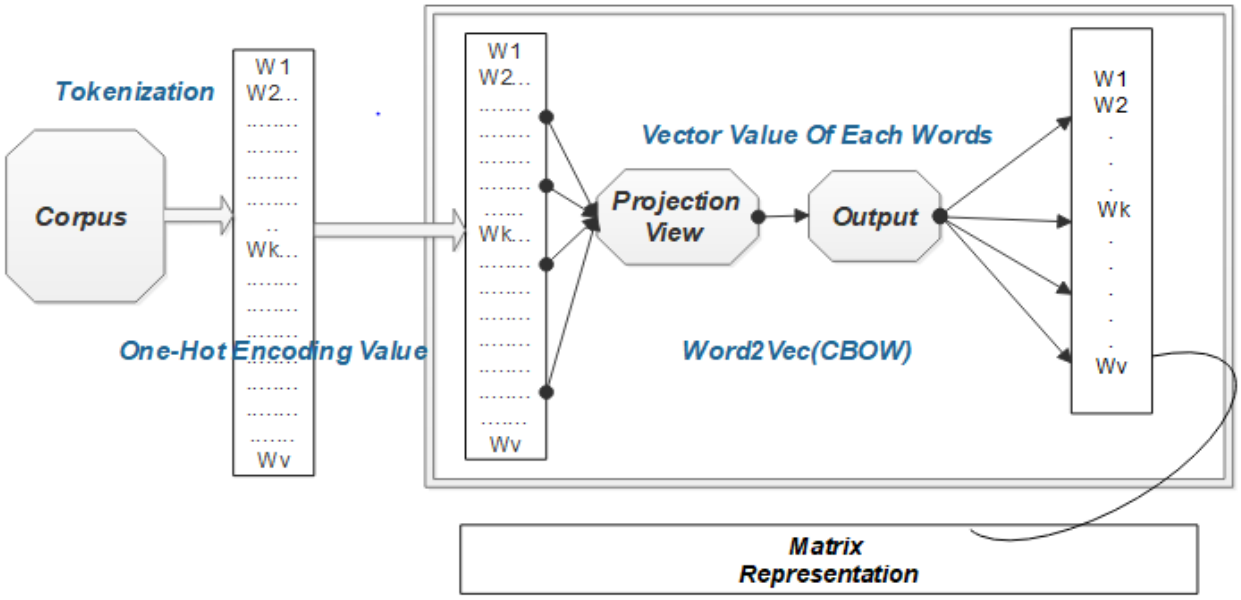

Figure 3. 3 Continuous Bag of Word model

Every word has been unique encoding values that can represent both for context words as input values to the neural network and target word as for finding the vector values. In figure 3.4 shows that, as taking the final rows for illustration. The encoding value [4, 1, 6, 7] is the context word ('ፊዴራላዊ', 'ስርዓት', 'ጥረቶችን', 'ሁሉ'), and [5] are the target word ('ለመታደግ'). The index value '4' for context word 'ፊዴራላዊ' represent at position 4 puts one otherwise zero with its vocabulary size, The index value '5' for the target word 'ለመታደግ' represent at position 5 puts one otherwise zero with its vocabulary size, and the other representation also given in these formatting.

```
[[4, 1]]
[3]
Context_word: ['PAD', 'PAD', 'ፌዴራላዊ', 'ስርዓት'] -> Target_word: የሀገሪቱን
[[3, 1, 5]]
[4]
Context_word: ['PAD', 'የሀገሪቱን', 'ስርዓት', 'ለመታደግ'] -> Target_word: ፌዴራላዊ
[[3, 4, 5, 6]]
[1]
Context_word: ['የሀገሪቱን', 'ፌዴራላዊ', 'ለመታደግ', 'ጥረቶችን'] -> Target_word: ስርዓት
[[4, 1, 6, 7]]
[5]
Context_word: ['ፌዴራላዊ', 'ስርዓት', 'ጥረቶችን', 'ሁሉ'] -> Target_word: ለመታደግ
```

Figure 3. 4 Encoding representation for context and target words

In figure 3.4 shows the encoding representation value for each word in the corpus. From these selected word2vec model, there is generated a vector value for each word. In these, the order of words is changed. As an increase in the number of window size, we can gain more contextual representation value for each word but there is creating a bottleneck in the network, it has taken a more computational power and also more generated times. Because of these, we have taken a window size of 2 for representing each word vector value. In figure 3.4 shows that the vector representation value of each word in the corpus.

	0	1	2	...	253	254	255
እንሆናለን	-0.228881	-0.003130	0.673105	...	0.099049	-0.506783	-0.127541
የሀገሪቱን	-0.402091	-0.297442	0.228016	...	0.209179	0.457892	0.274865
ፌዴራላዊ	-0.240298	0.230187	-0.234696	...	0.334411	-0.089914	0.037590
ለመታደግ	-0.005285	-0.182688	0.154544	...	0.170273	-0.095251	0.099494
ጥረቶችን	-0.385871	0.481563	0.057006	...	0.310794	0.062152	-0.246049

Figure 3. 5 Generated vector value for each words

3.2.3. Sentence embedding

In the word embedding translating each word to equivalent vector value, the next step is we have converted the utterance in vector value format. In the sentence embedding, we discussed some embedding techniques for converting each utterance into corresponding vector values, we have selected a mean embedding technique.

3.2.3.1. Mean embedding

We have selected the mean embedding techniques for taking the average weight value of each word that exists in the utterance. This embedding technique is more favorable than others because of saving a lot of time and computational power. As the utterance contains (W1, W2… Wn), we have tokenized the utterance into words that can fetch each word values from the representations of the word2vec model. The mean embedding can be represented in the equation (3.11).

$$Smean = \sum_{j=1}^{n} \frac{Wj}{n} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3.11)$$

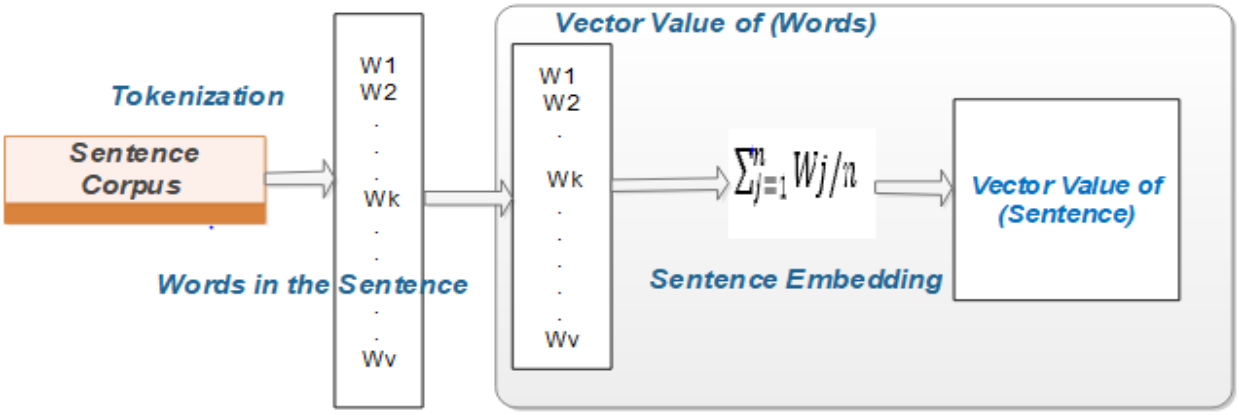

Figure 3. 6 Mean embedding model

In figure 3.4 shows that, the vector representation value of each word in the corpus. As the datasets can be prepared in utterance formats. We have performed a mean embedding technique to encode the meaning of the utterance that enables us to understand the linguistic context in which the word said. It is useful for predicting the value of the utterance, computing the similarity between utterances, and also for estimating the intention of the utterance. Figure 3.6 shows some of the utterance vector value with embedding size of 256.

```
['የሀገሪቱን', 'ፌዴራላዊ', 'ስርዓት', 'ለመታደግ', 'ጥረቶችን', 'ሁሉ', 'እንዲቀላቀሉ', 'ጥሪ', 'አቅርበናል']
[-0.04468092022222222, 0.06538011133333334, -0.06926335122222221, 0.048816687666666664, -0.088982760111111
['በማህበራዊ', 'ዘርፍም', 'በትምህርትና', 'ጤና', 'አገልግሎት', 'ተደራሽነት', 'መልካም', 'ውጤቶችን', 'አስመዝግቧል']
[-0.03417076022222223, 0.06366941399999998, -0.0951551618888889, 0.08201921566666666, -0.06388726833333333
['በኢትዮጵያን', 'ሕይወት', 'ደህንነቱ', 'የተጠበቀ', 'እንዲሆን', 'ከውጭ', 'እያደገ', 'የመጣውን', 'ስጋት', 'ማረም', 'አለብን']
[-0.08457176245454545, 0.06696645172727272, -0.10595656363636365, 0.08185132627272727, -0.0868003757272727
['የመድኃኒት', 'አቅርቦትና', 'ስርጭት', 'የገጠሩን', 'ህብረተሰብ', 'ማዕከል', 'ባደረገ', 'ሁኔታ', 'እንዲከናወን', 'ይደረጋል']
[-0.0901037681, 0.04616668859999999, -0.0758804987, 0.08964915339999999, -0.07342098789999998, 0.057910018
['ሚድያዎቻችንና', 'ጋዜጠኞቻችን', 'እዉነትን', 'ለማረጋገጥ', 'አይሰሩም']
[-0.007717055000000002, 0.038446117, -0.0698742148, 0.0464977854, -0.025266464799999992, 0.051677095799999
```

Figure 3. 7 Generated vector value for each utterance

These dense vector representation for each utterance are the preprocessing or gaining feature from the textual data in numerical format. These numerical format value of each utterance can be feed to the proposed CNN deep learning approach for both performing feature engineering and classification tasks.

3.2.4. Convolutional network approach

In the final component of the system model, we have been applied CNN architectures for both feature extraction and classification phases of the overall system scenario. In the feature extraction phase, we have converted each vector values of the utterance into a one-dimensional

array value (flatten). In the classification phase, we have used a different activation function for training the network and estimating its output values from the proposed network model.

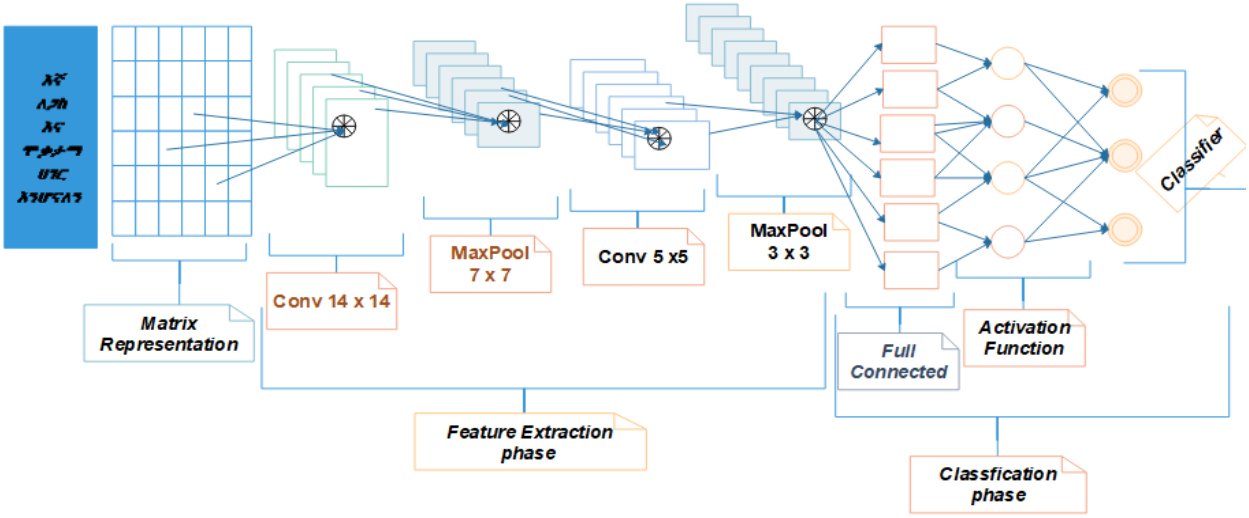

Figure 3. 8 Proposed CNN Architecture

3.2.4.1. Feature extraction

In the feature extraction phase, there are sentence embedding values of each utterance value. We have feed the matrix representation of utterance value to the convolutional neural network as a 16 x 16-dimensional matrix as input. For extracting features, we have used the four-layer of the convolutional neural network from convolutional layer to fully connected layer.

In the first convolutional layer, we have used a filter size with 3 x 3 with one filter value without specifying any stride and padding value to the input value. We have multiplied the 16 x 16-dimensional matrix by the filter size and adding the value we have gain 14 feature maps as 14 x 14-dimensional matrix value. This value has been passed to the ReLu function for converting the negative values from the feature map to zero and maximum value. After applying the convolve and ReLu function, next we have fed a 14 x 14-dimensional matrix to the pooling layer. In the first pooling layer, we have specified the stride and window size values to be 2 and the pooling approach have been preferred max pooling for this implementation, finally, we have performed the entire matrix representation we can gain a 7 x 7-dimensional matrix values.

In the second convolutional layer, we have also used a filter size with a 3 x 3 matrix with one filter values without specifying the values of stride and padding. We have accepted the 7 x 7-dimensional matrix then multiplied by a 3 x 3 filter size. From this, we have gained a 5 x 5

matrix representation as output values from the second convolutional layer. The values that come from the convolution layer can be pass to the ReLu function for removing negative value from the feature map. After gaining values from the ReLu function. In the pooling layer, we have used stride and window size value can be 2 and max-pooling approaches. From the pooling layer, we have gained a 3 x 3-dimensional matrix value. At the fully connected layer, we have gained a one-dimensional array of 9 scalar values. These nine values have been taken as an input (Xi) for representing one utterance value to the neural network.

3.2.4.2. Classification phase

In the feature extraction stage, we have been gaining a one-dimensional array list values for each utterance value. After that, we have specified the weight, bias, learning rate, batch size and other parameters that can be needed for the models. We have discussed the selected activation function in our architectures as bellows.

3.2.4.2.1. Scenario one (SoftMax function)

As put the processing scenarios for invoking the specified activation function for training the models and SoftMax as result classifier. In the initial state, we have declared the parameters for the models to train like weight, bias, learning rate, training input, training output, batch size, and the number of epoch for training and testing purposes. For performing this processes, we have followed some organized steps as follows.

In step one, we have accepted the training input and label value, the weight, and bias for reducing the noise data for the training and testing phase, batch size, and also the learning rate for adjusting the bias and weight value in the back propagation phase.

In step two, we have performed input value to weight matrix-like multiplication in the dense layers at the same time add a bias value to each neuron, and also a summation function in the dense layer. For estimating the output value from the dense layer and inputs to the next layer, we can feed to the specified transfer functions (ReLu, Softplus, Softsign, sigmoid and Swish).

In step three, we have accepted the output value from the previous layer as an input value and performed an input value to weight matrix-like multiplication at the same time adding bias value to each neuron. We have applied a summation of the values in each neuron for feeding the Softmax activation function for estimating the result values in the output layer.

In step four, we have calculated the Cross-Entropy Loss for reducing the loss from the model. We have also performed a stochastic gradient descent optimizer with momentum for adjusting the weight, bias values for training the model, and also to improve the model performance and reduce the error rates. Finally, we have performed each step repeatedly as the loss for each class value can be minimum value and gaining optimal weighted vale and as the accuracy level is the maximum value.

3.2.4.2.2. Scenario two (Sigmoid function)

As in applying the sigmoid activation function in the convolutional neural network in the output layer. In the initial stage, we have declared the weight, bias, learning rate, training input, output value, batch size, and the number of epoch for training and testing purposes of the model. For performing this functions, we have followed some organized steps.

In step one, we have performed the weight with input value multiplication at the same time adding bias values for each neuron, applying a summation function in the dense layer, and also giving this output value to the specified activation function (ReLu, Softplus, Swish)for estimating the output value in the dense layer.

In step two, we have accepted the output value of the previous layer as an input value. We have performed input value to weight matrix multiplication at the same time adding bias value for each neuron and also applying the summation function in each neuron. We used a sigmoid function in the output layer for estimating output values as predication.

In step three, for reducing the loss function and improving the prediction performance of the model, we have calculated the binary cross entropy loss. We have used a stochastic gradient descent optimizer with momentum for reducing the error rate, updating the initial weight, and bias in the back propagation step to improve the performance of the model.

In step four, we have performed each step repeatedly as the loss for each class value can be minimum value and as the accuracy level is the maximum value.

CHAPTER FOUR

4. EXPERIMENTAL RESULT AND SIMULATION SETUP

4.1. Introduction

In this chapter, we have discussed the general point of view experimental setup and simulation results of the proposed system for Amharic political speech utterance. In the first section, we have been discussed the speech act and intent data sets that have been used for training and testing purposes for the proposed model. In the second section, we have described the simulation environment and tool for implementing the system as viewing its evaluation purpose. In the third section, we have discussed the experimental configuration and the parameters to be used in the configuration scenario. Finally, we have discussed the results that can be evaluated by the performance metrics.

4.2. Data set

The corpus data set have been collected from the Amharic political party manifesto of its document publication for reading purposes like (Ethiopian Democratic Party (EDP, 1997 and 2007 E.C.), Ethiopian people revolutionary party (EPRP, 2001 E.C.), Kinijit Manifesto (1997 E.C.), Ethiopian People's Revolutionary Democratic Front (EPRDF, 2007 E.C.), Ethiopian prosperity party (2012 E.C.). The other data sets collected from the websites like (

www.zehabesha.com/amharic/, www.addisadmassnews.com/, from each part official social media pages) that every politician can suggest your own ideas for the stated information about your own program or give some guideline or contradictory idea for group ideology.

As the corpus data sets have been collected, we have selected each information feeding ideology from each document for speech act and intention sense utterance. For annotated each utterance into the corresponding intentional sense value of utterance, it can require a professional. In this study, we have asked two independent professionals in the analysis of political speech in different sectors and the identification of each speech based on the word uttered in the utterance. In the case of literature professionals, we have been gained the analysis of each utterance, and also the identification of one utterance from the other utterance depends on the action verbs exist and the transmitted information to the concerned body in that utterance. For justifying or classifying one utterance intention type, there is an action verb in that utterance. In the case of the political science department profession, we have gained how to select style-oriented service, politician opinion, politician attitude, and consumer-oriented activity utterance from the documented file. We have been filtering 2,975 utterances for the three speech acts and their intentional type.

	Corpus size	Vocabulary size
Size	2, 975 (utterance)	17, 733 (word)

Table 4. 1 The data sets

The analysis and preparation of datasets based on the professionals gaining ideology and giving is the target point of directions. We have discussed some illustration below:

Utterance	The intent express in U	Speech intent	Speech act
እኛ ለጋስ እና ሞቃታማ ሀገር እንሆናለን።	ሀገርን ለጋስ እና ምቹ ማድረግ	Promise (achieving for future dream)	Commissive
ሚድያዎቻችንና ጋዜጠኞቻችን እዉነትን ለማረጋገጥ አይሰሩም።	በትክክል እንደማይሰሩ ማሳወቅ ወይም የሚሰሩትን ስራ መቃወም እና መተቸት	Claim (you are not doing well and criticize your work)	Assertive
የሀገሪቱን ፊዴራላዊ	ያለውን ነገር ከግብ ለማድረስ	Request (Ask others for ideas, abilities, or	Directive

ስርዓት ለመታደግ ጥረቶችን ሁሉ እንዲቀላቀሉ ጥሪ አቅርበናል።	ከሌሎች የሀሳብ፣ አቅም ወይም የእውቀት ፍላጎት መጠየቅ	knowledge to achieve their goal)	
የኢትዮጵያን ሕይወት ደህንነቱ የተጠበቀ እንዲሆን ከውጭ እያደገ የመጣውን ስጋት ማረም አለብን።	ያሉት ስጋቶች እንዲስተክከሉ የሀሳብ ምንጭ ወይም ምክረ ሀሳብ ማቅረብ	Advice (Provide a source of advice or advice to address existing concerns)	Directive

Table 4. 2 Prepare and analysis the utterance act and intention type

We have assigned each utterance into the speech act and intent class for labeling each utterance by considering the action verb that can be expressed in the utterance. The utterance can express for doing something in the feature and the user can desire to gain something in the coming generation for intention type of promise.

	Assertive	Commissive	Directive
Size	1, 318 (utterance)	648 (utterance)	1, 009 (utterance)

Table 4. 3 Data sets of speech act utterance

As depending on the speech act, we have labeled the intent type of each utterance as generalizing the class of act utterance value. Firstly, we have labeled the assertive act utterance into the report and claim intention type based on the action verb used in the utterance. In the report, we have been express the speaker to be described, disclose, announce, make a known, issue about the state, express, narrate, and also advertise the transmitted information to the audience. The claim is also, we have been express the speaker can be a protest, argue, defend, testify, objection, complaint, disapproval, disagreement, opposition, dissent for once transmitted information by other group members or politicians party.

Secondly, we have labeled the commissive act utterance into the promise and offer intention type based on the action verb used in the utterance. In the promise, we have been express the speaker can be a pledge, swear, covenant, solemnly promise, and give hope of the transmit information for the audience to do something in the future. In the offer, we have been express the speaker can be provided, give, supply, donate, and contribute for something to the public.

Finally, we have labeled the advice, order, and request intention types based on the action verbs used in the directive act utterance. In the advice, we have been express the speaker can give counsel, recommend, guidance, hints, tips, directions and offer suggestions about the transmitted information by other group members or political parties. In the order, we have been express the speaker can give a command, enjoin, decree, rule, mandate, and determine for once state information. In the request, we have been express the speaker can be solicited, appeal for, requisition, and implore the information to the other group members or politician.

	Claim	Report	Offer	Promise	Advice	Order	Request
Size	383	935	215	433	711	136	160

Table 4. 4 Number of utterance for intention type

As increasing the number of data sets, we have applied a synonym replacement approach for fabricating other data sets. First, we have selected an action verbs from the corpus that can produce an illocutionary force for that utterance. Based on the verbs exist in the utterance we have applied a cosine similarity for finding other synonym verb in the corpus. The original utterance 2975 is after augmenting we can gaining anther 2975 utterance as a total of 5950 utterances can be gaining. Finally, we have to gain the utterance vector values for training and testing purposes, and also divided the data set into 70% for training and 30% for testing the model evaluation. In the table **4.5** shows the utterance datasets distribution for each speech intent to act value distribution.

Speech act	Assertive		Commisive		Directive		
Size of utterance	2636		1296		2018		
Intent type	Claim	Report	Offer	Promise	Advice	Order	Request
Size of utterance	766	1870	430	866	1422	272	360

Table 4. 5 Datasets after data augmentation

4.3. Simulation environment

In this study, for encode, process, and yield the result from the proposed system architecture as for the collected data. We have to use a window operating system with corei5, CPU processor as RAM size of 8GB for preparing the documentation, and also encoding the coding parts. We have

used a python programming language for implementing the proposed model. Python 3.7.0 version of programming language can integrate with the anaconda environments for accessing the freely available python libraries and packages. The anaconda environment can provide a python built-in package and library.

As tools, we have used the Keras library on the backend of the Tensorflow library for preprocessing the raw text into a numerical representation. The Keras library is an open-source and also operating in the CPU for windows operating system. The Keras library can provide a word2vec package. As python code editors, we can use a PyCharm editor of 2018.

4.4. Experimental scenario

To identify, examine, and investigating the effect of activation function for training and testing of the model, we have considered two scenarios in this study. In the first scenario, there is an input neuron value accept from the fully connected layer, we have taken Xavier random weight initialization technique for compensating vanishing gradient problem as the model becomes deeper. After performing a multiplication input neuron value with weight in the dense layer, we have been applied an activation function (ReLu, Softplus, Softsign, sigmoid, and Swish) for estimating the dense layer neuron values. After gaining the dense layer neuron value, the next step is taken the multiplication of dense layer neuron value to its weighted value, for estimating its predicated value in the output value, we have been taken a Softmax activation function for normalized its output value and also applied a cross-entropy loss.

As the model performance can converge, optimal weighted value, and minimizing loss functions for reducing the difference between predicted and actual value, we have applied an optimization algorithm. In this case, we have used a stochastic gradient descent with momentum with gamma value ***0.0001*** for accelerated the training process. We have used an epoch ***10*** and batch size value ***64*** for training and testing of the model. For compensating its weighted value and training process faster as taking a learning rate value ***0.1*** have been taken in this procedure.

In the second scenario, there is an input neuron value accept from the fully connected layer, we have taken Xavier random weight initialization technique for compensating vanishing gradient problem as the model becomes deeper. After performing a multiplication input neuron value with weight in the dense layer, we have been applied an activation function (ReLu, Softplus, and

Swish) for estimating the dense layer neuron values. After gaining the dense layer neuron value, the next step is taken the multiplication of dense layer neuron value to its weighted value, for estimating its predicated value in the output value, we have been taken a sigmoid activation function. As the model performance can converge, an optimal weighted value gained, and minimizing loss function for reducing the difference between predicted and actual value, we have applied an optimization algorithm. In this case, we have used a stochastic gradient descent with momentum as gamma value ***0.0001*** for accelerated the training process. We have used an epoch ***10*** and batch size value ***64*** for training and testing of the model. For compensating its weighted value and the training process faster learning rate value ***0.1*** have been taken in this procedure.

4.5. Experimental result and discussion

In this study, the experimental result has been viewed in two scenarios depending on the type of activation function for invoking in the CNN architecture and also as used in the output layer. Based on the stated evaluation metrics we have been shown each activation function achievement. In the first scenario, we have discussed the SoftMax classifier results with some activation functions (Softsign, Softplus, ReLu, sigmoid, and Swish) for training the model by applied the stated performance metrics, in the second scenario we have discussed the sigmoid activation function in the output layer results with some activation functions (Softplus, ReLu, Swish) for training the model by applied the stated performance metrics for the Amharic speech act and intent classification tasks.

4.5.1. Experimental result of scenario one

In this section, we have shown the proposed Amharic speech act and intent classification experimental results by applying the SoftMax classifier. We have also visualized the experimental result for both training and testing results using accuracy measurement. We have used a confusion matrix to visualizing the correct and incorrect label value of utterance from the model. And also statistics matrix performance report for each class's values.

In table ***4.6*** shows the classification performance accuracy value of the model. The utterance datasets have been organized in superclass, subclass, and subclass to superclass interdependence format as the model to be train and test phase. In the first case, we have taken the superclass

(speech act value of utterance like assertive, commissive, and directive) as train the model at the same time we have performed testing phase for knowing performance accuracy.

In the second case, we have taken the subclass (speech intent value of utterance with 7 class) as train the model at the same time we have performed the testing phase for knowing performance accuracy. In the third case, we have taken the interdependence (hierarchical format) of the speech intent to act corresponding, in this case first, we have train the model the intent utterance value to act utterance. In the testing phase, first, we have checked the utterance are correctly predicted to the learned intent value, after that we have checked the correctly predicted intent utterance value traced to the corresponding utterance speech act value.

No	Activation function		Speech Act	Speech Intent	Intent based speech act (hierarchical) format.
	Dense layer	Output layer			
1	Softsign	Softmax	92.5%	89.2%	83.8%
2	Sigmoid		88.0%	80.9%	81.6%
3	Softplus		84.8%	71.4%	71.6%
4	Swish		80.5%	64.2%	64.5%
5	ReLu		80.5%	64.2%	64.5%

Table 4. 6 Softmax classifier performance with different activation function

In this experimental results, we have discussed each of the activation function performance value achievement strategically behavior for this proposed model. As taking the Softsign, we have achieved a better performance result than others. Due to, handling the vanishing gradient problem in the back propagation stage of the network, maintain the value of neuron greater than zero for the parameters can be updated its initial value for preserving the neuron values to be lost and also model converges to optimal result, for tackling this problem we have serves the model performance values optimal solution for the speech act and intent classification phase. We have achieved a performance testing accuracy value of 92.5%, 89.2% and 83.8% for speech act, speech intent and the intent based speech act classification respectively.

As taking the sigmoid activation function we have been achieved better result next to Softsign. This result can be achieved due to the data preprocessing of each utterance imbalanced class distribution value for the sigmoid function perform better, as a usual the specified parameter value to be passed there is no occur vanishing gradient or the neuron cannot lost its value, and

also this function can specify output value range between 0 and 1, there are no encountered stack overflow errors in the output layer activation function, from this regarded we have been achieved performance testing accuracy value 88.0%, 80.9% and 81.6% for speech act, speech intent and the intent based speech act classification respectively.

As taking the Softplus, we have achieved good result next to sigmoid. This function can be put in the third rank performance accuracy results of the model. As for tackling in the back propagation stage for updating the initial parameter values of the model, there are neuron values cannot be lost due to the coverage value of both negative and positive results. And the neuron values too high the model have depressed by vanishing gradient problem, this problems can be tackled due to the derivative value of this functions converges to one. As for compensating this unexpected problems that can be faced from parameter values or this utterance dataset, we have achieved better result next to sigmoid function as performance testing accuracy value 84.3%, 71.4% and 71.6% for speech act, speech intent and the intent based speech act classification respectively.

As for taking the swish and ReLu activation functions, as the number of class increase in this scenario both function results have yield minimum compare from others. The swish and ReLu function uses to train the model, these two functions have been produced too high values feed the Softmax classifier, this value encountered a stack overflow error in the output layer. This problem leads to the model cannot converge to optimal results. As taking ReLu in the back propagation stage cannot handle vanishing gradient problems and also swish partially vanishing problems can exist, because of these and other parameter value variables, the weight value of utterance has led to almost less performance value have been achieved from the model. Both of the two activation functions achieve the same performance testing accuracy value of 80.5%, 64.2% and 64.5% for the speech act, speech intent and the intent based speech act classification respectively.

In table ***4.6*** shows that we have performed the SoftMax classifier with different activation functions. In the stated results, the Softsign activation function has been trained the models and also applied as a Softmax classifier with batch size value of 64 for train the models, we have achieved a better performance result from the other activation functions, from these we have selected this activation function for train the model.

4.5.1.1. Performance analysis for speech act classification using SoftMax classifier

In this section, we have to visualize the training and testing performance accuracy values from the above table ***4.6*** shows the Softsign activation function has been achieved better performance results from the other activation function, from this we have selected the Softsign functions to visualize the results in graphical format.

In figure ***4.1*** shows the number of epoch for train and test the model increases, the performance value of the model also increase consistency. The performance accuracy value of the model for both training and testing accuracy have been higher after one epoch. As the number of epoch increase after the first epoch the accuracy value have increased slightly. Due to, the algorithm can allow a back propagation in the network for reducing the cost function (difference between predicated and actual value) and obtaining an optimal weighted value for compensating the weighted gradient and vanishing gradient problem in model for achieving better results and also the parameter value used in the model like bath size value of 64 sampled dataset have been used before the model update. as the optimization algorithm, stochastic gradient decent algorithm with momentum for obtaining optimal weighted, bias value and reducing cost function as taking a parameter value of learning rate 0.1 and gamma 0.0001. We have achieved the training accuracy of the model is 94.9% and the testing accuracy of the model is 92.5%, by applied a Softsign activation function.

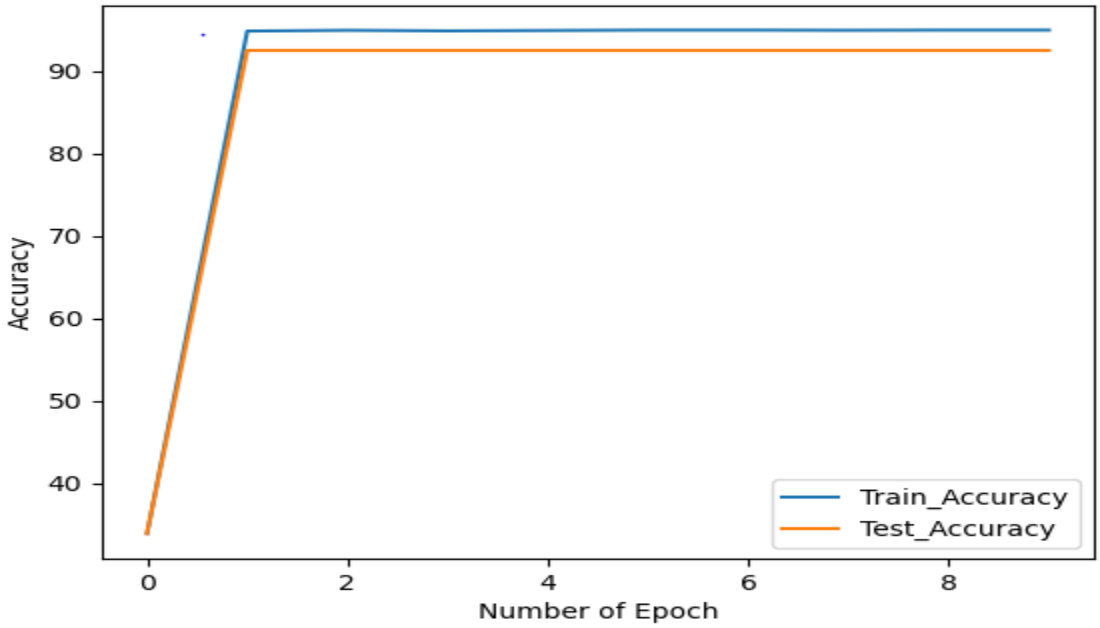

Figure 4. 1 Performance accuracy for speech act classification using SoftMax classifier

As the confusion matrix figure shows that the speech act testing accuracy with predicted and actual values of each class accuracy in diagrammatical format. In this case, we have used 30% (1,788) of data sets for testing purposes. As shows the diagram assertive class (790) values the correct labels are (747) and also for incorrect label value for directive (21) and (22) value for commissive, as a directive class (606) for correct labels is (562) and also for incorrect label value for assertive is (23) and (21) for commissive, and finally, as a commissive class (392) for the correct label (345) and for incorrect label value for assertive is (28) and (19) for the directive. This shows that we have achieved better results after training the model to predicate unknown data. In figure ***4.2*** shows this results in diagrammatical formats as a true label with predicate labels, data distributions.

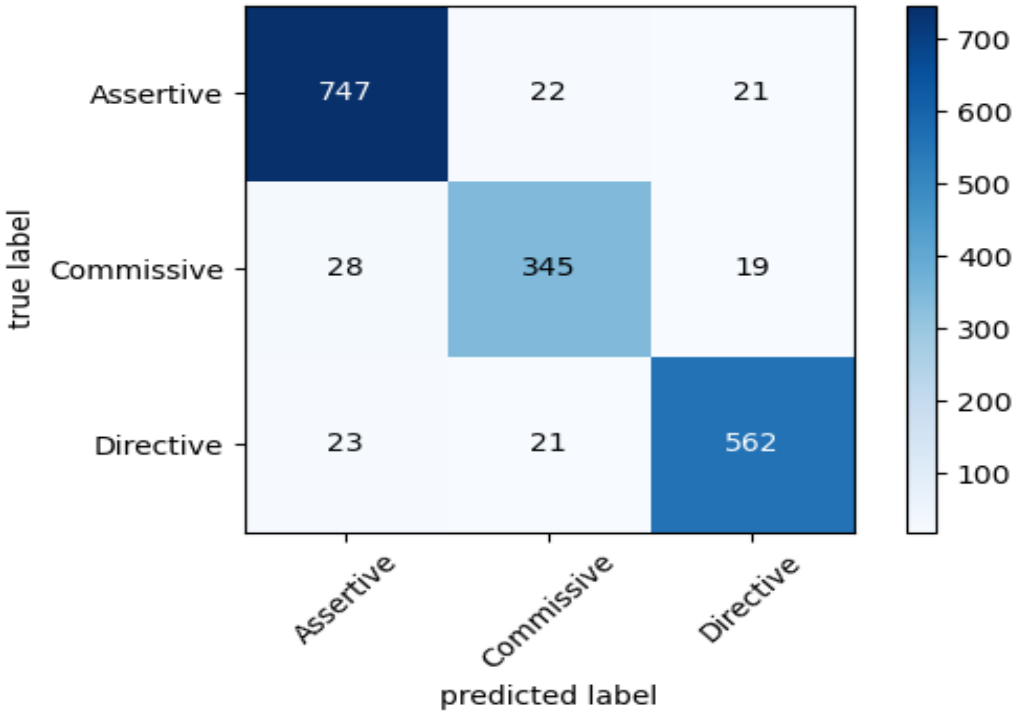

Figure 4. 2 Confusion matrix for speech act classification using SoftMax classifier

In this section, we have to visualize the confusion matrix values for each class distributions as the completeness, exactness, and true values of the model prediction performance. We have used the three statistics measurement report mechanism like recall, precision, and F measure for each class. In table ***4.7*** shows the detailed performance values of each class. In this case, the performance of majority class (assertive and directive) is higher than the minority class (commissive). As for training the model with majority class value, in the case of tested the model classifies the more data we have learned. This shows that the classification accuracy of the minority class is dominated by the majority class.

	Assertive	Commissive	Directive
Precision	93.6%	88.9%	93.4%
Recall	94.6%	88.0%	92.7%
F measure	94.1%	88.5%	93.0%

Table 4. 7 Performance report for speech act classification using SoftMax classifier

In this section, we have to visualize the loss values during the training and testing phase of the model. In the SoftMax classifier neural network, the cross-entropy loss measure values are good measurement technique. We have used a categorical cross-entropy loss for predicting the loss

values in the training and testing accuracy for the speech act multi-classification tasks probability distribution of its utterance.

As viewed in figure ***4.3***, the number of epoch increase consistency for measuring the performance of training and testing accuracy the loss values also decrease consistency. As stated above the model exceed at epoch accuracy value almost constant at the same time loss values from the model are also constant.

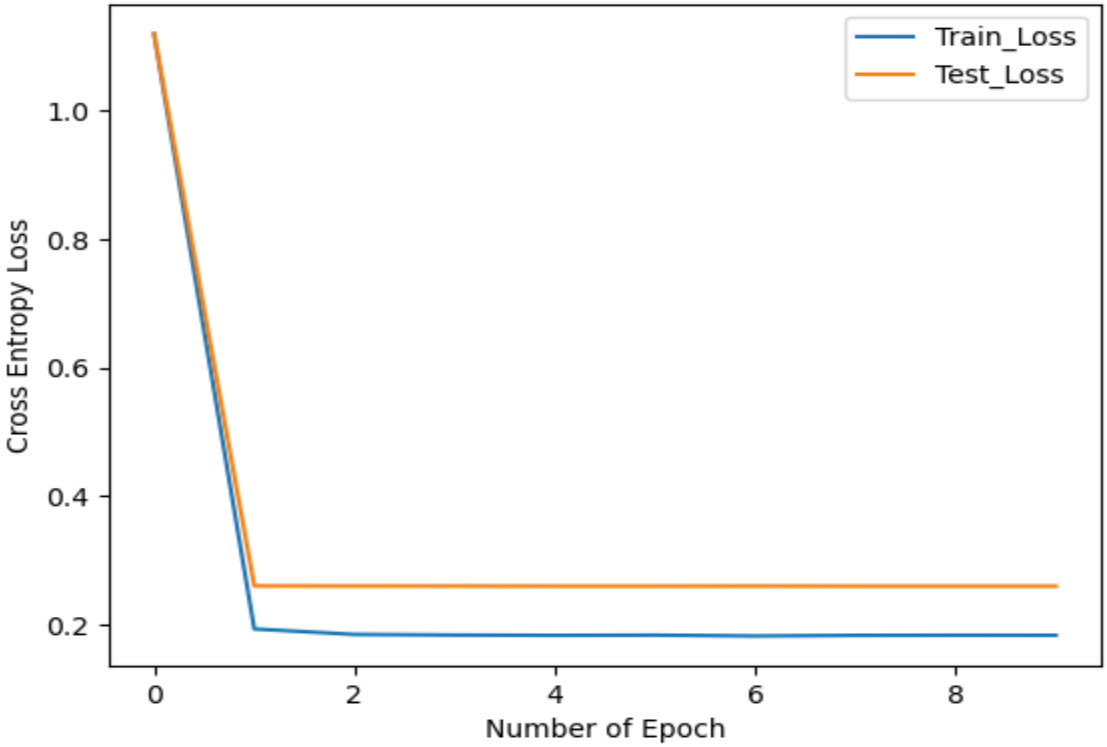

Figure 4. 3 Cross entropy loss for speech act classification SoftMax classifier

4.5.1.2. Performance analysis for speech intent classification using SoftMax classifier

In section, we have shown the accuracy value, the confusion matrix of each intent class distribution to view the incorrect and correct label of utterance, and also the statistics measurement report of each intent class. As expressed in the above table **4.6,** the Softsign functions have been achieved better performance result for utterance intent classification task.

As figure **4.4** shows, the number of epoch for training and testing of the model increase the performance value of both training, and testing accuracy also increases consistency. The performance accuracy value of the model for both training and testing accuracy have been higher

after one epoch. As the number of epoch increase after the first epoch the accuracy value have increased slightly. Due to, the algorithm can allow a back propagation in the network for reducing the cost function (difference between predicated and actual value) and obtaining an optimal weighted value for compensating the weighted gradient and vanishing gradient problem in model for achieving better results and also the parameter value used in the model like bath size value of 64 sampled dataset have been used before the model update. as the optimization algorithm, stochastic gradient decent algorithm with momentum for obtaining optimal weighted, bias value and reducing cost function as taking a parameter value of learning rate 0.1 and gamma 0.0001. We have measured the performance values as knowing data and unknown data value for the model. Finally, we have achieved a training accuracy of 93.0% and testing accuracy of 89.3% value.

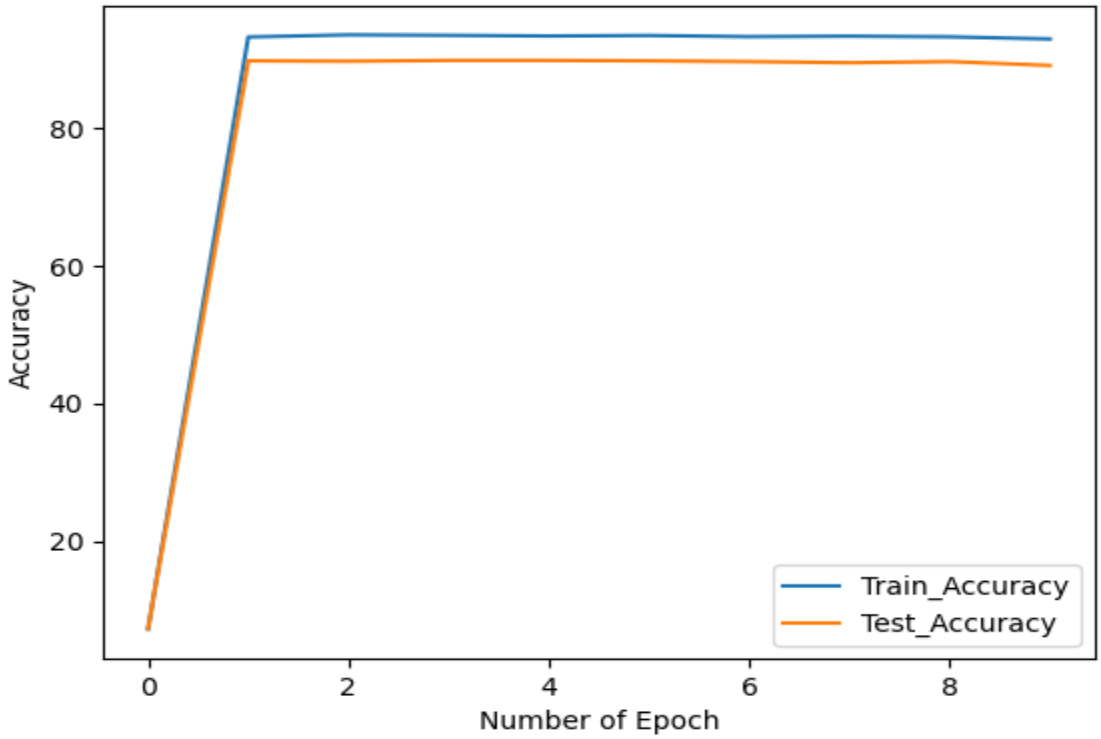

Figure 4. 4 Accuracy for speech intent classification using SoftMax classifier

As a view, the testing accuracy in diagrammatical view we have used a confusion matrix for using 30% (1,788) of the data sets for the report (561), claim (230), promise (261), offer (130), advice (427), order (82), and also request (97) value distribution. As the result showed that, the number of classes can increase the measuring value for each class value less. Because the utterance for assigned for each class value not proportional to each other. We have applied a

SoftMax at the output layers with this number of class the performance values can be more favorable than others used Softsign functions by handling the vanishing gradient problem. In figure ***4.5*** shows, the test accuracy values for each class values in numerical format as estimating each utterance values to correct values.

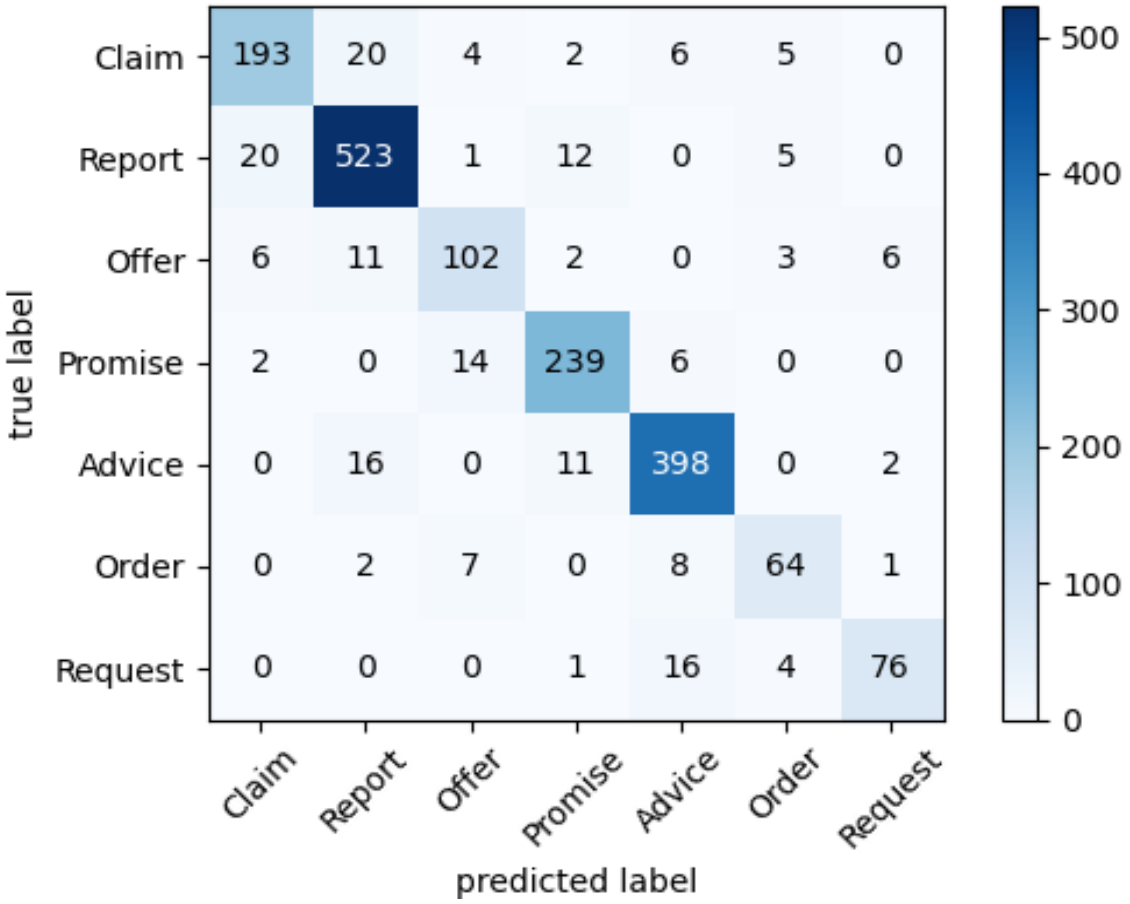

Figure 4. 5 Confusion matrix for speech intent classification using SoftMax classifier

In this section, we have to visualize the confusion matrix values for each intent class distributions as the completeness, exactness, and true values of the model prediction accuracy. We have used the three statistics measurement report mechanism like recall, precision, and F measure for each class.

In table ***4.8*** shows the detailed performance values of each class. In this case, the minority class (offer, order, and request) are less utterance distribution than others, due to these the classification performance values of minority class are dominant by majority class. In the request class, the recall values are less, due to the model classify its true value to other class. In the case

of order and offer class the recall and precision performance values are less, due to the model classify its true value to other majority class and also the other majority class true values are incorrectly classified as order or offer class.

	Precision	Recall	F measure
Claim	87.3%	83.9%	85.6%
Report	91.4%	93.2%	92.3%
Offer	79.7%	78.5%	79.1%
Promise	89.5%	91.6%	90.5%
Advice	91.7%	93.6%	92.5%
Order	79.0%	78.0%	78.5%
Request	89.4%	78.4%	83.5%

Table 4. 8 Performance report for speech intent classification using SoftMax classifier

In this section, we have to visualize the loss values during the training and testing phase of the model. In the SoftMax classifier neural network, the cross-entropy loss measure values are good measurement techniques. We have used a categorical cross-entropy loss for predicting the loss values in the training and testing accuracy for the speech act multi-classification tasks probability distribution of its intent utterance.

As viewed in figure ***4.6***, the number of epoch increase consistency for measuring the performance of training and testing accuracy the loss values also decrease consistency. As stated above the model exceed at epoch two accuracy value almost constant at the same time loss values from the model are also constant.

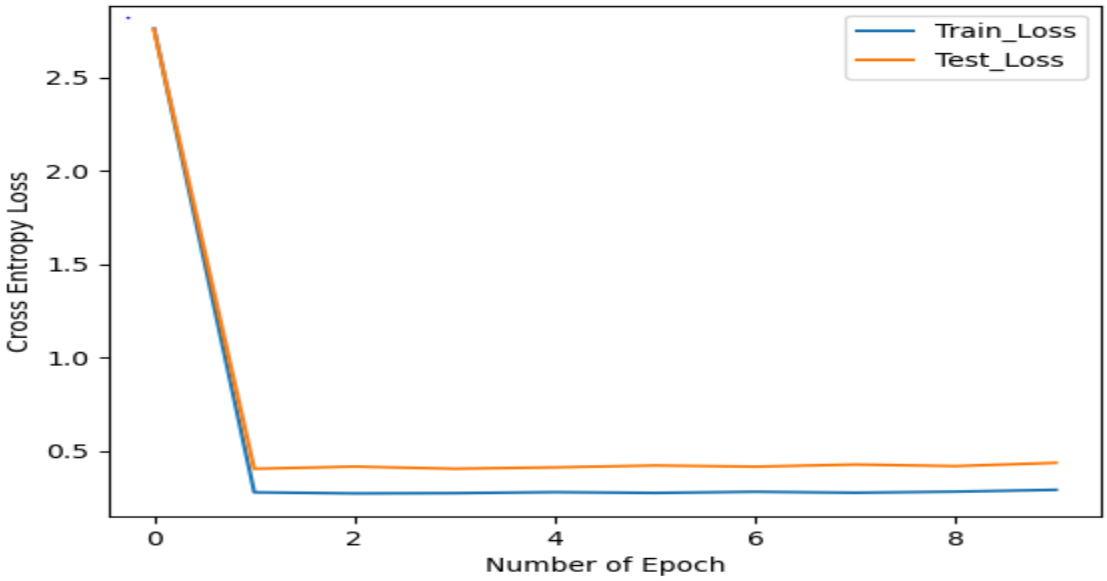

Figure 4. 6 Cross entropy loss for speech intent classification using SoftMax classifier

4.5.1.3. Performance analysis for speech act and intent classification using SoftMax classifier

In this section, we have been discussed the interdependence of speech act and intent classifications as a bottom-up approach. To visualizing this performance results of the model, we have only expressed the estimation accuracy value of the model. To be showing better results in the above table ***4.6***, the Softsign function has been achieved better results from the others. In this processes, first, we have checked the utterance is correctly classified as the correct trained model intent values. After that, we have checked the intent values with the corresponding speech act values as a bottom-up approach structure. In figure **4.7**, shows the intent to act interdependence in graph format with train and test accuracy of 89.0% and 83.2% respectively. The performance accuracy value of the model for both training and testing accuracy have been higher after one epoch. As the number of epoch increase after the first epoch the accuracy value have increased slightly. Due to, the algorithm can allow a back propagation in the network for reducing the cost function (difference between predicated and actual value) and obtaining an optimal weighted value for compensating the weighted gradient and vanishing gradient problem in model for achieving better results and also the parameter value used in the model like bath size value of **64** sampled dataset have been used before the model update. as the optimization algorithm, stochastic gradient decent algorithm with momentum for obtaining optimal weighted, bias value and reducing cost function as taking a parameter value of learning rate **0.1** and gamma **0.0001**.

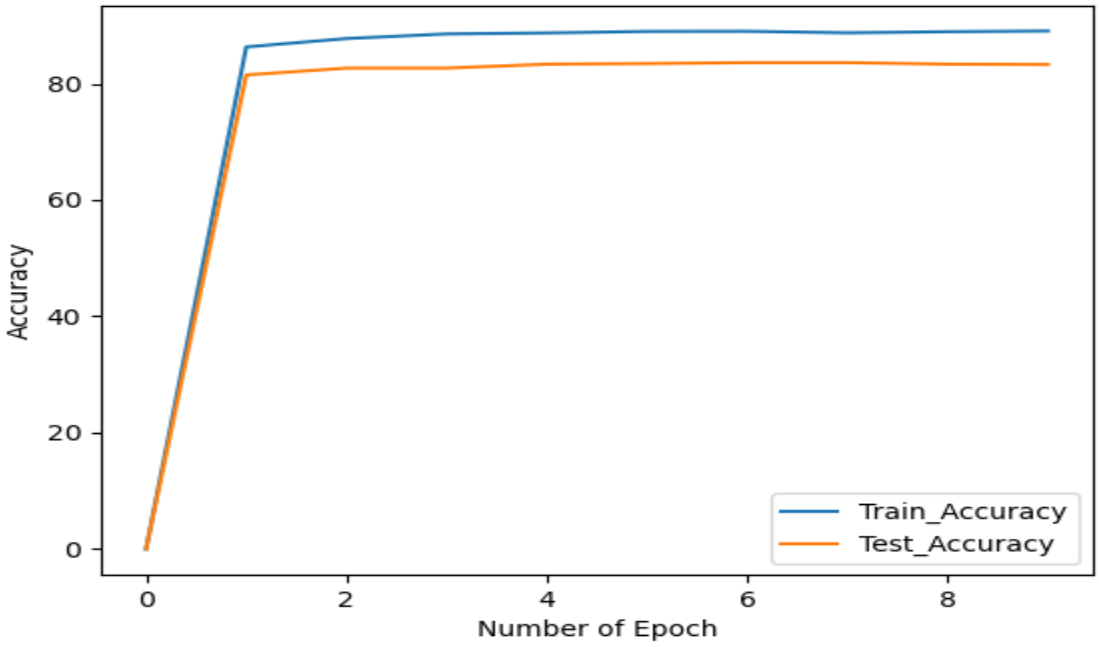

Figure 4. 7 Accuracy for speech act and intent classification using SoftMax classifier

In this section, we have to visualize the loss values during the training and testing phase of the model. In the SoftMax classifier neural network, the cross-entropy loss measure values are good measurement technique. We have used a categorical cross-entropy loss for predicting the loss values in the training and testing accuracy for the speech act multi-classification tasks probability distribution of its intent utterance. As viewing in the above diagram shows the number of epoch increases performance value also increases, as coming to figure **4.8** shows that the loss value from the model also decrease. This show that the models have been exceeded in the matured stage for learning from the utterance data sets speech intent to act corresponding situations.

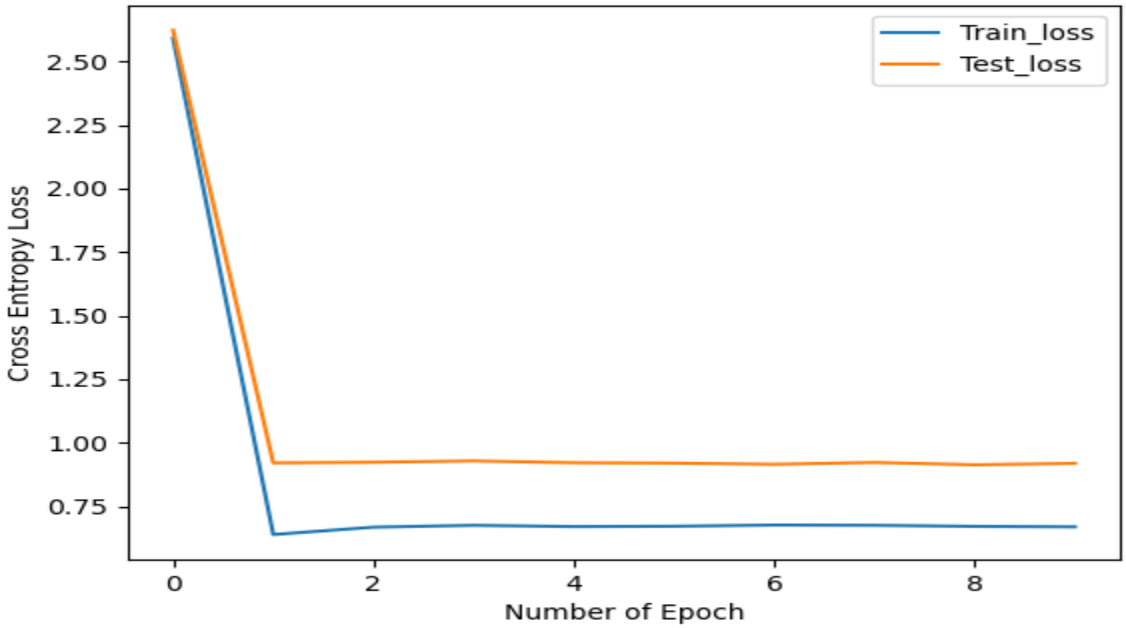

Figure 4. 8 Cross entropy loss for speech act and intent classification using SoftMax classifier

4.5.2. Experimental result for scenario two

In this section, we have discussed the experimental result of the proposed Amharic speech act and intent classification tasks. As we selected the activation function (ReLu, Softplus, and swish) to train the model and also used the sigmoid function for estimating the output value in the output layer of the neural network. As for evaluating the proposed model, we have used an accuracy performance measurement for the training and testing phase of the model, a confusion matrix for visualizing the testing accuracy value of correct and incorrect label values in diagrammatical format and also a statistics matrix performance report for visualizing the testing accuracy value of each utterance label value.

In table ***4.9*** shows the classification performance accuracy value of the model. The utterance datasets have been organized in superclass, subclass, and subclass to superclass interdependence format as the model to be train and test phase. In the first case, we have taken the superclass (speech act value of utterance like assertive, commissive and directive) as train the model at the same time we have performed testing phase for knowing performance accuracy.

In the second case, we have taken the subclass (speech intent value of utterance with 7 class) as train the model at the same time we have performed the testing phase for knowing performance

accuracy. In the third case, we have taken the interdependence (hierarchical format) of the speech intent to act corresponding, in this case first, we have train the model the intent utterance value to act utterance. In the testing phase, first, we have checked the utterance are correctly predicted to the learned intent value, after that we have checked the correctly predicted intent utterance value traced to the corresponding utterance speech act value.

No	Activation function		Speech Act	Speech Intent	Intent based speech act (hierarchical) format.
	Dense layer	Output layer			
1	ReLu	Sigmoid	88.1%	58.6%	56.8%
2	Softplus		88.0%	59.6%	50.1%
3	Swish		87.7%	57.7%	54.4%

Table 4. 9 Sigmoid function performance value with different activation function

In this experimental results, we have discussed each of the activation function performance value achievement strategically behavior for this proposed model. As we have used the ReLu function to train the model, it has achieved a better performance result from the other activation functions as a sigmoid function in the output layer. These performance values have been achieved because of the data preprocess weighted value of each utterance or the specified parameter value, this condition can handle the vanishing gradient problem or there is no happened a neuron value lost during the train and testing phase of the model and also the class imbalanced distribution of each utterance happened, this serves the sigmoid function can classify the more data to learn each speech act. In the speech intent classification because of the number of class enlarge, the class imbalanced distribution of each utterance, and each intent utterance values have been mutually exclusive property the ReLu function in the dense layer and sigmoid function in the output layer achieved less results than other functions. From this, we have been achieved a performance testing accuracy value of 88.1%, 58.6%, and 56.8% for the speech act, speech intent, and the intent-based speech act classification respectively.

As used the Softplus, this activation function has been put in the second rank of result for speech act and the first rank of result for speech intent classification from the other activation function. Because of the data preprocess utterance weight value, specified parameter value or class imbalanced distribution for each speech intent, as for solved the vanishing gradient problem, as

preventing neuron value lost for train the model by converging both negative and positive values. From this, we have achieved better performance results for speech intent classification than other activation functions. The performance testing accuracy values are 88.0%, 59.6%, and 50.1% for the speech act, speech intent, and the intent-based speech act classification respectively.

As used the swish, this activation function is an improvement of the ReLu function as for resolving the problems that occurred in ReLu. In this case, we have been achieved reverse results. As because of the model network structure, data preprocess weight value, the specified parameter value, the class imbalanced distribution of each speech utterance, or the inter-correlation between swish function in the dense layer with sigmoid function in the output layer. We have achieved a performance testing accuracy value of 87.7%, 57.7%, and 54.4% for the speech act, speech intent, and the intent-based speech act classification respectively.

The table ***4.9*** shows that we have used a learning rate 0.1, a batch size value 64 with epoch 10 for train the model, and estimating the performance values of each activation function. In the stated results, the ReLu activation function has been trained the model and also applied as a sigmoid function in the output layer, we have achieved a better performance result from the other activation functions, from this we have selected this activation function for train the model. In the first case, we have analyzed the speech act classification by applied the performance measure, in the second case, we have analyzed the speech intent classification by applied the performance measure and also in the third case we have shown the interdependence between speech intent to act (hierarchical) format classification.

4.5.2.1. Performance analysis for speech act classification using sigmoid function

In this section, we have to visualize the training and testing performance accuracy values from the above table ***4.9*** shows the ReLu function has been achieved better performance results from the other activation functions, from these we have selected the ReLu functions to visualize the results in graphical format.

In figure ***4.9*** shows, the number of epoch for train and test the model increases, the performance value of the model also increase consistency. The performance accuracy value of the model for both training and testing accuracy have been higher after one epoch. As the number of epoch increase after the first epoch the accuracy value have increased slightly. Due to, the algorithm

can allow a back propagation in the network for reducing the cost function (difference between predicated and actual value) and obtaining an optimal weighted value for compensating the weighted gradient and vanishing gradient problem in model for achieving better results and also the parameter value used in the model like bath size value of 64 sampled dataset have been used before the model update. as the optimization algorithm, stochastic gradient decent algorithm with momentum for obtaining optimal weighted, bias value and reducing cost function as taking a parameter value of learning rate 0.1 and gamma 0.0001. As the model reached to the specified epoch range, we have achieved the training accuracy of the model is 90.5% and the testing accuracy of the model is 88.1%, by applied a ReLu function in the dense layer of the model.

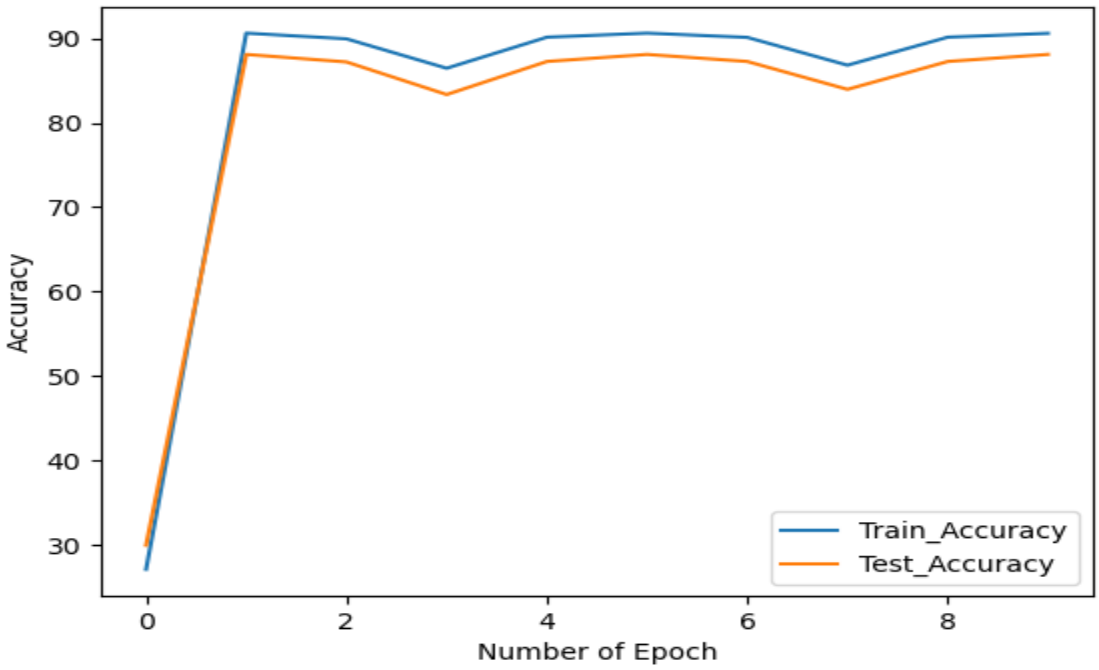

Figure 4. 9 Accuracy for speech act classification using sigmoid function

As the confusion matrix figure shows that the speech act testing accuracy with predicted and actual values of each class accuracy in diagrammatical format. In this case, we have used 30% (1,788) of data sets for testing purposes. As shows the diagram assertive class (790) values the correct labels are (790) and also for incorrect label value for a directive (0) and (0) value for commissive, as a directive class (606) for correct labels is (530) and also for incorrect label value for assertive is (63) and (13) for commissive, and finally, as a commissive class (392) for the correct label (255) and for incorrect label value for assertive is (46) and (91) for the directive.

This show that we have achieved better results after training the model to predicate unknown data.

The above figure ***4.9***, shows this results in diagrammatical formats as a true label with predicate labels for the utterance distributions to each speech act class. This can give general accuracy values as models are a distribution manner. In figure ***4.10***, represented each class performance values. From this as the number of datasets for one class has been enlarged the performance value also better as applying the sigmoid function in the output layer. The assertive utterance values have dominated from the other two utterance value, from this purely classified to the specified utterance value.

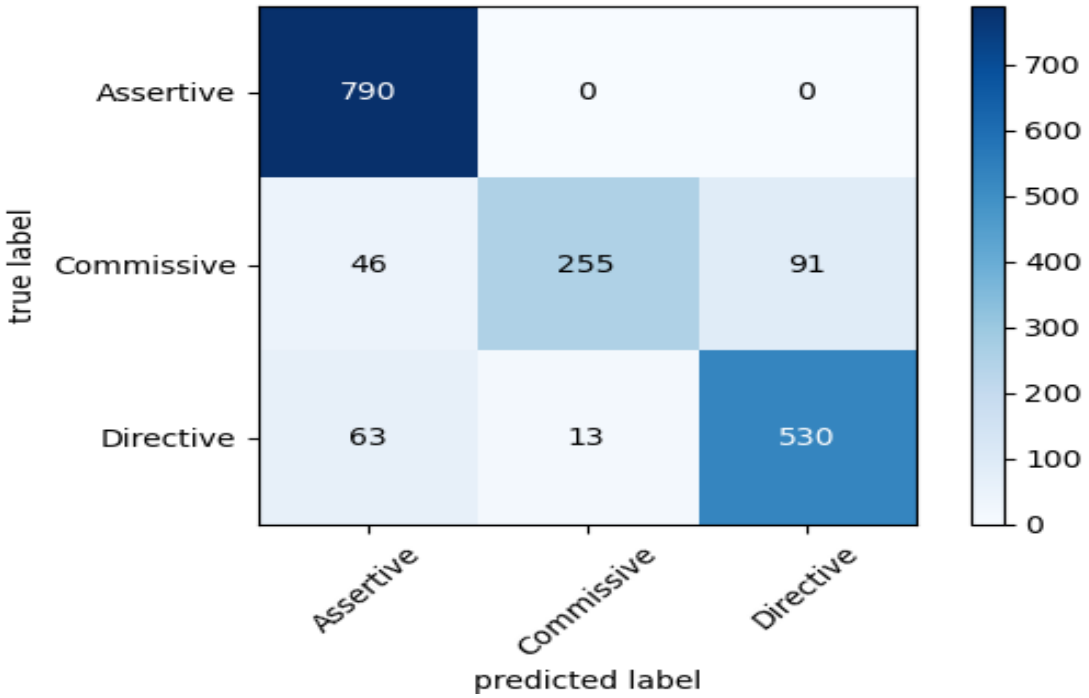

Figure 4. 10 Confusion matrix for speech act classification using sigmoid function

In this section, we have to visualize the confusion matrix values for each class distributions as the completeness, exactness, and true values of the model prediction accuracy. We have used the three statistics measurement report mechanism like recall, precision, and F measure for each class. In table ***4.10*** shows the detailed accuracy values of each class. The minority class (commissive) recall values are less from the other two classes, due to the sigmoid functions are classified as the more learned utterance of its majority class. In this case, the true label value of its minority class is classified to the majority class.

	Assertive	Commissive	Directive
Precision	87.9%	95.1%	85.3%
Recall	100%	65.1%	87.5%
F measure	93.5%	77.3%	86.4%

Table 4. 10 Performance report for speech act classification using sigmoid function

In this section, we have to visualize the loss values during the training and testing phase of the model. In the sigmoid function in the neural network, the binary cross-entropy loss measure values are good measurement techniques for multiple classification tasks. We have used a binary cross-entropy loss for predicting the loss values in the training and testing accuracy for the speech act multi-classification tasks probability distribution of its utterance. As viewed in the figure ***4.11***, the number of epoch increase consistency for measuring the performance of training and testing accuracy the loss values also decrease consistency. As stated above the model exceed at epoch two accuracy value almost increase slowly at the same time loss values from the model are also decrease slowly.

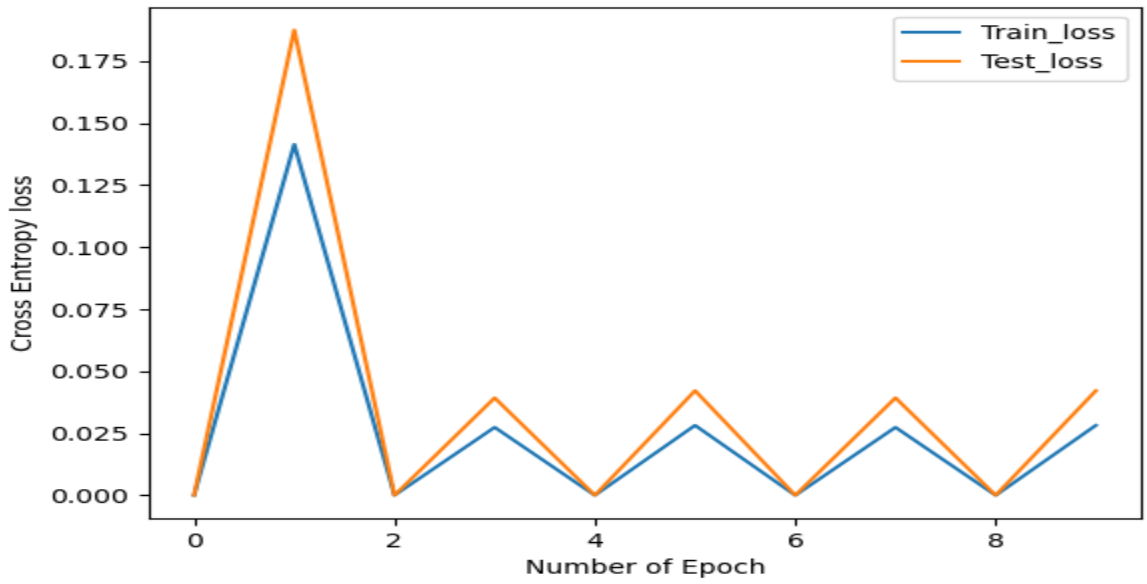

Figure 4. 11 Cross entropy loss for speech act classification using sigmoid function

4.5.2.2. Performance analysis for speech intent classification using sigmoid function

In this section, we have analyzed the speech intent classification by applied a sigmoid function in the output layer. As expressed in the above table ***4.9***, the Softplus functions have been achieved

better performance results for utterance intent classification task. As figure ***4.12*** shows, the number of epoch for training and testing of the model increase the performance value of both training, and testing accuracy also increases consistency. As the number of epoch to train the model to exceed 2, the performance value of the model almost increases a little, this shows that the model can be mature after epoch 2. We have used a learning rate value 0.1 and ten epoch for trained the model. We have measured the performance values as knowing data and unknown data value for the model. Finally, we have achieved a training accuracy of 66.3% and testing accuracy of 59.6% value.

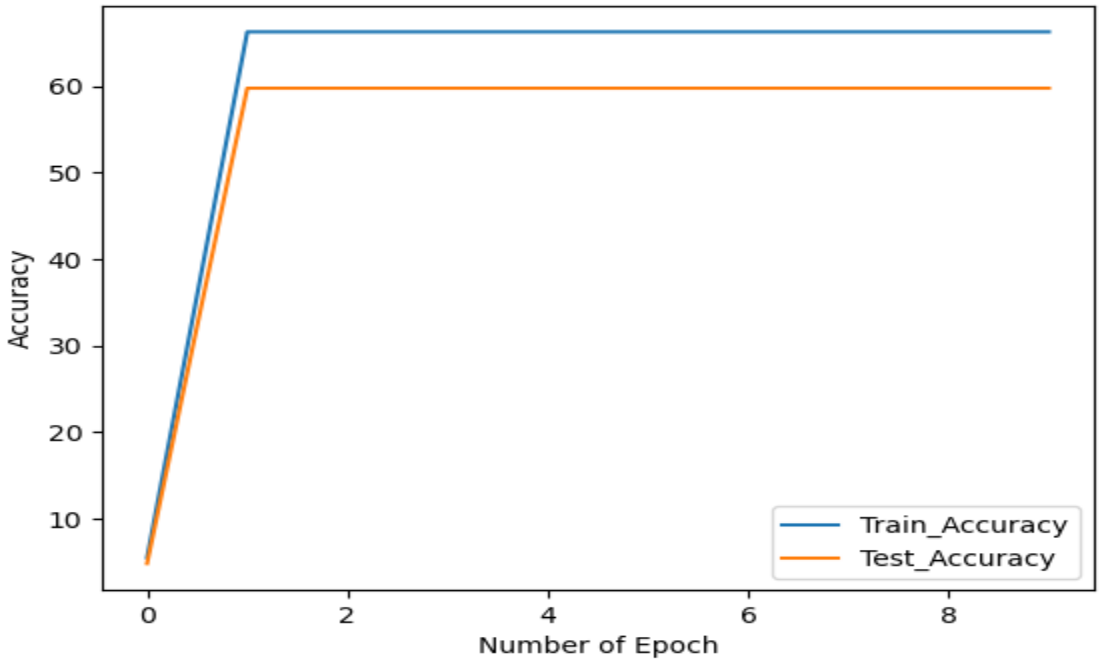

Figure 4. 12 Accuracy for speech intent classification using sigmoid function

As viewing the result, the number of classes of this utterance increases the performance accuracy values by using sigmoid function are minimum compared to the softmax classifier results. This shows that the sigmoid function for multiple classifications is not feasible for this scenario.

In this section, we have to visualize the loss values during the training and testing phase of the model. The sigmoid function in neural network, the binary cross-entropy loss measure values are good measurement techniques for multiple classification tasks. We have used a binary cross-entropy loss for predicting the loss values in the training and testing accuracy for the speech intent multi-classification tasks probability distribution of its utterance.

As viewed in the figure ***4.13***, the number of epoch increase consistency for measuring the performance of training and testing accuracy the loss values of the model also decrease consistency. The model shows that the test loss values are less than train loss values, from this as the number of data sets for training the models to increase the train loss value decrease. As using sigmoid function in the output layer for multi classification task the number of utterance for each labels values also increase.

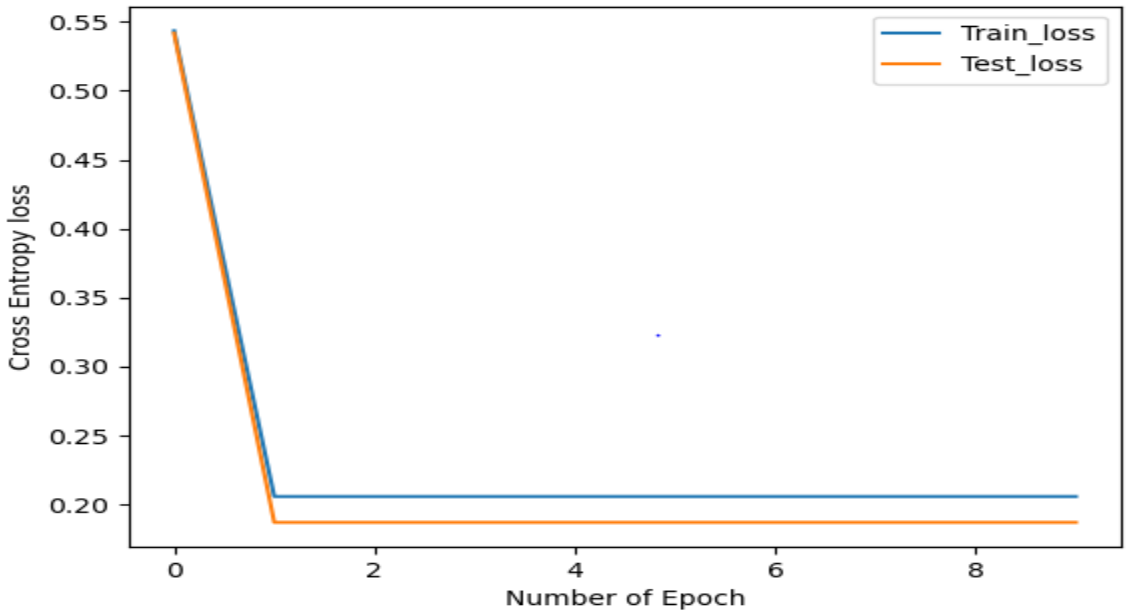

Figure 4. 13 Cross entropy loss for speech intent classification using sigmoid function

4.5.2.3. Performance analysis for speech act and intent classification using sigmoid function

In this section, we have been discussed the interdependence of speech act and intent classifications as a bottom-up approach. To visualizing this performance results of the model, we have only expressed the estimation accuracy value of the model. To be showing better results in the above table ***4.9,*** the ReLu functions have been achieved better results from the Softplus, swish function.

In this processes, first, we have checked the utterance is correctly classified as the correct trained model intent values. After that, we have checked the intent values with the corresponding speech act values as a bottom-up approach structure. As for performance measure for cross-checking both the speech act and intent values for one utterance as a training accuracy 62.7% and testing

accuracy 54.4%. In figure ***4.14***, shows the speech intent to act interdependence in graph format of train and test accuracy.

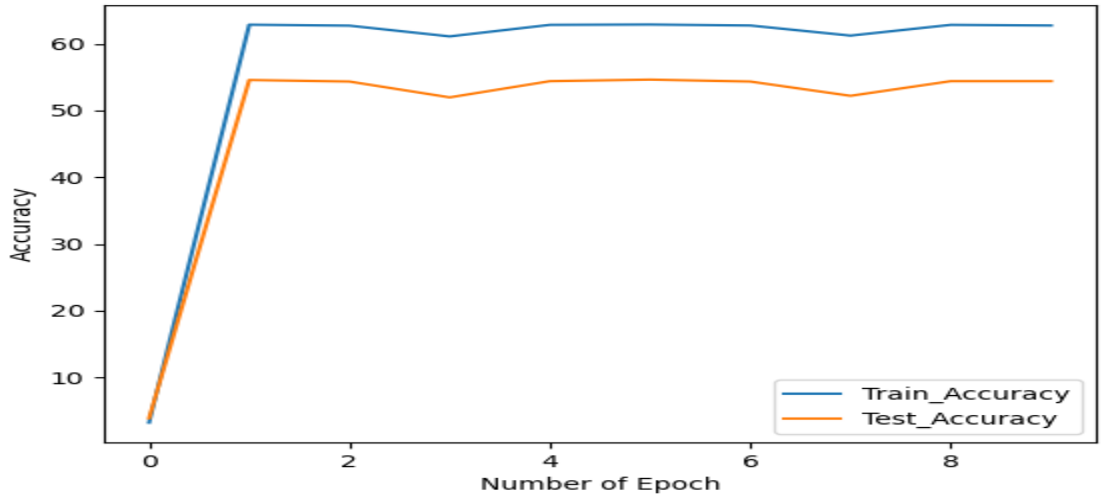

Figure 4. 14 Accuracy for speech act and intent classification using sigmoid function

In this section, we have to visualize the loss values during the training and testing phase of the model. The sigmoid function in the output layer for multi-classification process of a neural network, the binary cross-entropy loss measure values are good measurement technique. We have used a binary cross-entropy loss for predicting the loss values in the training and testing accuracy for the speech intent to act multi-classification tasks probability distribution of its utterance. As viewing in the above diagram shows the number of epoch increases the performance value also increases, as coming to figure ***4.15*** shows that the loss value from the model also decrease. This show that the models have been exceeded in the matured stage for learning from the utterance data sets of speech intent to act interdependence situation.

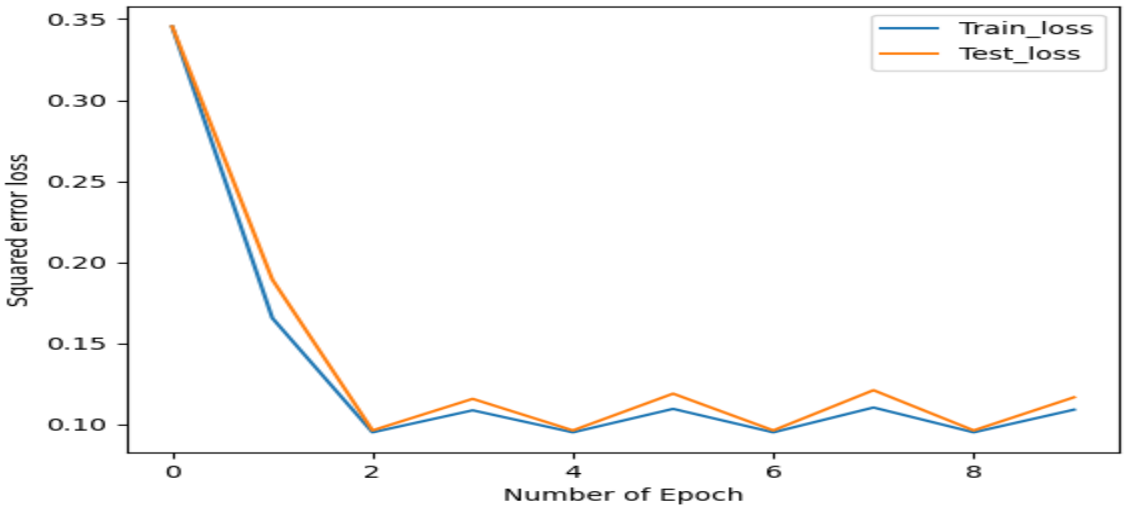

Figure 4. 15 Cross entropy loss for speech act and intent classification using sigmoid function

For analysis, the interdependence between speech act and intent classification hierarchal format is not feasible, due to the intended value for each utterance have mutually exclusive. As applying the sigmoid function for the CNN architecture performance values can be diverge

CHAPTER FIVE

5. CONCLUSION AND FUTURE WORK

5.1. Conclusion

In this study, we have presented the Amharic political speech act and intent identifications. To evaluate the model we have used a new data set of speech acts and intent identification tasks for Amharic political speech. In the process, we have used a word2vec over the sentence embedding models for preparing the utterance, and for predicting the speech act and intent, we used the Softsign, swish, Softplus, ReLu, sigmoid and Softmax activation functions over the convolutional neural network.

In the experiment, we have taken two scenarios for identifying the speech act and intent. In the first scenario, we took a Softmax result classifier with the Softsign to train the model and achieved a better performance value of 92.5%, 89.3%, and 83.8% for the speech act, intent, and intent to act classification respectively. In the second scenario, we took a sigmoid result classifier with the ReLu to train the model and achieved a better performance value of 88.1% and 56.8% for the speech act and intent to act classification, and also the Softpus to train the model performance values are 59.6% for the speech intent classification.

We conclude that from the experiment result in the proposed model has achieved a better performance result with this class imbalanced training data sets and the parameter for used in the model to train and test as a classifier of the SoftMax with Softsign activation function.

5.2. Contribution

In this study, we have to develop a word embedding techniques with a deep learning approach of CNN for classifying the speech act and intent of Amharic political speech. The following have some contributions in this study.

- We have annotated a new speech act and intent data sets for analyzing and identifying once politician ideology.
- We have been conducted a combination of word embedding to sentence embedding data preprocessing for extracting better utterance datasets to reduce computational time and also achieve better performance values that have been gain..
- We have conducted an SLU utterance with the CNN algorithm for the bottom-up users speech intent to act generalized classification tasks have an optimal solution with a huge amount of data and also the utterance labels are mutually exclusive.
- We have conducted, which activation functions are suitable for this class imbalanced dataset for train and test the model and also result from classifier activation functions.

5.3. Future work

In the study, we have shown the speech act and intent classification task by applying a word embedding techniques with CNN for classification purpose. We have an additional task to improve the performance of speech act and intent classification for Amharic political speech some of them stated below.

- In the Amharic language for identifying the speech act and intent, there is no well-defined corpus that is tagged by professionals. In this study, we have collected small amount of data to enlarge the data, apply data augmentation techniques and also manually label the data sets. In the feature, for the model train and test purpose with a huge amount of data to inspect the error rate value and improve the performance rate.
- As conducting this study with other word embedding techniques (Glove, Fast text or Elmo) as feature extraction with a CNN classification with huge amount of data sets have been exist, can be improve the classification performance.
- As improving the intent to act (bottom-up) or act to intent (top-down) hierarchal structure classification performance. Once utterance value in the transmitted information have been depends on the pervious utterance value, for inspecting this and getting better performance accuracy conducting other deep learning approach.

REFERENCES

Austin , L. J. (1962). *How to do Things with Words.* Oxford University Press: Oxford.

Abien , F. A. (2019). Deep Learning using Rectified Linear Units (ReLU). *Neural and Evolutionary Computing*, 1-8.

Alom, M. Z., Tarek, T. M., & al, e. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *Computer Vision and Pattern Recognition*, 1 - 39.

Andrew , M., Raymond, D., & et al. (2011). Learning Word Vectors for Sentiment Analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologie. 1*, pp. 142-150. Portland, Oregon, USA: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/P11-1015

Atelach, A. A., & Lars, A. (2007). An Amharic Stemmer : Reducing Words to their Citation Forms. *Proceedings of the 5th Workshop on Important Unresolved Matters* (pp. 104–110). Prague, Czech Republic: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/W07-0814

Birol, K., Cuneyt, A., & Selman, D. (2019). An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing. *Association for Computing Machinery*, 1 - 4. doi:https://doi.org/10.1145/3342827.3342828

Chandra, K. D., & Afiahayati. (2018). Suitable CNN Weight Initialization and Activation Function for Javanese Vowels Classification. *INNS Conference on Big Data and Deep Learning 2018* (pp. 124–132). Yogyakarta, Indonesia: ELSEVIER. doi:https://doi.org/10.1016/j.procs.2018.10.512

David, G., Ben , A., Wei , L., & et al. (2006). A Closer Look at Skip-gram Modelling. *In Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)* (pp. 1222 - 1225). Genoa, Italy: European Language Resource Assocation (ELRA).

David, M. (2016). *How exactly does word2vec work?*

Demeke, G. A. (2010). The Origin of Amharic. *LINCOM Studies in Afrosiatic Linguistics, 28*, 99-110.

Dhana , K. S. (2019). Data Analytics: Role of Activation function In Neural Net. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(5), 299 - 302.

Dshahid. (2020, March 21). *towardsdatascience*. Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529

Faiz, A. S. (2017). Speech Communication. *Researchgate*, 1-10.

Fatih, E., & Galip, A. (2017). Data Classification with Deep Learning using Tensorflow. *2nd International Conference on Computer Science and Engineering* (pp. 755 - 758). IEEE.

François, C. (2018). *Deep Learning with Python.* United state of America: Manning Publications.

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 119-130.

Gao., H., Zhuang., L., Laurens., M., and Kilian., Q. (2016). Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*.

Gavrilov, A., Jordache, A., Vasdani, M., & Deng, J. (2018). Preventing Model Overfitting and Underfitting in Convolutional Neural Network. *International Journal of Software Science and Computational Intelligence*, 19-28.

Gokhan , T., & Renato, M. D. (2011). *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech.* india: Wiley.

Gokhan, T., & Renato, D. M. (2011). Systems for Extracting Semantic Information from Speech. In T. Gokhan, & D. M. Renato, *Spoken Language Understanding.* Noida, India: John Wiley & Sons, Ltd.

Hao , Z., Zhanlei , Y., & et al. (2015). Improving deep neural networks using softplus units. *in International Joint Conference on Neural Networks (IJCNN)* (pp. 1 - 4). IEEE. doi:10.1109/IJCNN.2015.7280459

James , O. S., Zuhair, B., & Keeley , C. (2010). A Machine Learning Approach to Speech Act Classification Using Function Words. *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications Agent and Multi-Agent Systems: Technologies and Applications* (pp. 82-91). Berlin, Heidelberg: Springer. doi:https://doi.org/10.1007/978-3-642-13541-5_9

Jason, B. (2020, July 31). *A Gentle Introduction to the Rectified Linear Unit (ReLU)*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

Jason, W., & Kai, Z. (2019). Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 6382–6388). Hong Kong, China: Association for Computational Linguistics.

Jeffrey, P., Richard, S., & Christopher, M. (2014). GloVe: Global Vectors for Word Representation. *In Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Jerry., W. (2020, 03 16). *Towards Data Science*. Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951

Jey, L. H., & Timothy, B. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *Proceedings of the 1st Workshop on Representation Learning for NLP*, (pp. 78–86). Berlin, Germany.

John, M. (2016). An Overview of Convolutional Neural Network Architectures for Deep Learning. *Microway, Inc*.

Junmei, Z., & William, L. (2019). Predicting Customer Call Intent by Analyzing Phone Call Transcripts Based on CNN for Multi-class Classification. *8th International Conference on Soft Computing, Artificial Intelligence and Applications* (pp. 9-20). Computer Science & Information Technology. doi:10.5121/csit.2019.90702

Kent , B. (2018). Speech acts and Pragmatics. In D. Michael , & H. Richard , *The Blackwell Guide to the Philosophy of Language* (pp. 147 - 167). San Francisco State University.

Konstantin , H. (2018). Data Augmentation and Deep Learning for Hate Speech Detection. *IMPERIAL COLLEGE LONDON*, 1 -96.

Kyoungman, B., & Youngjoong, K. (2018). Speech-Act Classification Using Convolutional Neural Network and Word Embedding. *International Journal on Artificial Intelligence Tools, 27*(06), 1-12. doi:https://doi.org/10.1142/S0218213018500264

Larsson., Gustav., Michael., M., and Gregory., S. (2016). FractalNet: Ultra-Deep Neural Networks without Residuals. *arXiv preprint arXiv:1605.07648*.

Lemon, O. (2011). Learning What to Say and How to Say it: Joint Optimisation of Spoken Dialogue Management and Natural Language Generation. *Computer Speech and Language, 25*(2), 210–221. doi: http://dx.doi.org/10.1016/j.csl.2010.04.005

McTear, M. F. (2004). *Spoken dialogue technology. Toward the conversational user interface.* Springer. doi:http://dx.doi.org/10.1007/978-0-85729-414-2

Mikolov, T., & Quoc, L. V. (2014). Distributed Representations of Sentences and Documents. *compuatation and language*. Retrieved from http://arxiv.org/abs/1405.4053

Mikolov, T., Kai , C., & et al. (2013). Efficient Estimation of Word Representations in Vector Space. *Computation and Language*, 1 - 12.

Mridul, K. M., & Jaydeep, V. (2019). *Survey of Sentence Embedding Methods.* doi:10.13140/RG.2.2.21861.45289

Nikolay , K., Alexander, D., & Alexey , M. (2017). Development of a Model to Predict Intention Using Deep Learning. *National Research University Higher School of Economics*, 69-78.

Otis, H. G. (1969). Intentions and speech acts. *Analysis, 29*(3), 109–112. doi:https://doi.org/10.1093/analys/29.3.109

Pablo , F.-V. (2014). And Yet It Moves: The Effect of Election Platforms on Party Policy Images. *Comparative Political Studies, 47*(14), 1919–1944. doi:https://doi.org/10.1177/0010414013516067

Peng, J., Yue, Z., Xingyuan, C., & Yunqing, X. (2016). Bag-of-Embeddings for Text Classification. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, (pp. 2824 - 2830).

Piotr, B., Edouard, G., Armand, J., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. (S. Hinrich, Ed.) *Transactions of the Association for Computational Linguistics, 5*, 135–146. doi:http://arxiv.org/abs/1607.04606

Prabhu. (2018, March 4). *Understanding of Convolutional Neural Network (CNN) — Deep Learning*. Retrieved from Medium: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

Prajit, R., Barret , Z., & Quoc, V. L. (2017). Swish: A Self-Gated Activation Function. *Neural and Evolutionary Computing*, 1 -12.

Rabiner, L. R. and Huang, B .H. (1993). *Fundamentals of Speech Recognition.* Englewood eciffe.NJ.

Radhika , K., Bindu , K. R., & Latha , P. (2018). A Text Classification Model Using Convolution Neural Network and Recurrent Neural Network. *International Journal of Pure and Applied Mathematics, 119*(15), 1549-1554. Retrieved from http://www.acadpubl.eu/hub/

Ramón, L.-C., Zoraida, C., & et al. (2014). Review of spoken dialogue systems. *Loquens, 1*(2). doi:doi: http://dx.doi.org/10.3989/loquens.2014.012

Renato , M. D. (2013). Spoken Language Understanding: A Survey. *IEEE Signal Processing Magazine*.

Rikiya, Y., Mizuho, N., Richard, K. G., & Kaori, T. (2018). Convolutional neural networks: an overview and application in radiology. *Springer, 9*, 612-629. doi:https://doi.org/10.1007/s13244-018-0639-9

Ryan, K., Yukun, Z., & et al. (2015). Skip-Thought Vectors. *Computation and Language*, 1 -11.

Searle, J. R. (2012). *Speech Acts.* Cambridge: Cambridge University Press. doi:https://doi.org/10.1017/CBO9781139173438

Shivashankar, S., Trevor, C., & Timothy , B. (2019). Target Based Speech Act Classification in Political Campaign Text. *Association for Computational Linguistics* (pp. 273–282). Minneapolis: Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM).

Stefan, M. (2018). Prospective and retrospective rhetoric: A new dimension of party competition and campaign strategies. *In Manifesto Corpus Conference*.

Suhair, M. (2015). Speech Acts in Political Speech. *Journal of Modern Education Review*, *5*(7), 699–706. doi:10.15341/jmer(2155-7993)/07.05.2015/008

Tatsuya, H., Hiroyuki, S., & et al. (2019). Stochastic Tokenization with a Language Model for Neural Text Classification. *Conference: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1620–1629). Florence, Italy: Association for Computational Linguistics. doi:10.18653/v1/P19-1158

Víctor, L. S., Eduardo , E., & et al. (2012). A Case Based Reasoning Model for Multilingual Language Generation in Dialogues. *Expert Systems with Applications, 39*(8), 7330–7337. doi:http://dx.doi.org/10.1016/j.eswa.2012.01.085

Victor, Z., & Stephanie, S. (2008). Spoken Dialogue Systems. In B. Jacob, S. Mohan , & A. H. Yiteng , *Speech Recognition* (pp. 705-722). Berlin, Heidelberg: Springer. doi:https://doi.org/10.1007/978-3-540-49127-9_35

Wei, P., Kainan, P., & et al. (2018). Deep Voice 3: Scaling Text-To-Speech with Convolutional Sequence Learning. *Computer Science and Sound*.

Xavier, G., & Yoshua, B. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 249–256). Sardinia, Italy: Proceedings of Machine Learning Research. Retrieved from http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf?hc_location=ufi

Xu, G., Lee, H., Koo, M., and Seo, J. (2017). Convolutional Neural Network using a Threshold Predictor for Multi-label Speech Act Classification. *IEEE*.

Yoo, D., Ko, Y., & Seo, J. (2017). Speech-Act Classification Using a Convolutional Neural Network Based on POS Tag and Dependency-Relation Bigram Embedding. *IEICE Transactions on Information and Systems, 100-D*(12), 3081 - 3084. doi: 10.1587/transinf.2017EDL8083

Yue, G., Xinyu, L., Shuhong, C., & et al. (2017). Speech Intention Classification with Multimodal Deep Learning. *Advance Artificial intellegence*, 260–271.