Cape Peninsula University of Technology

ABSTRACTIVE TEXT SUMMARISATION USING RECURRENT NEURAL NETWORKS AT THE PARAGRAPH LEVEL

by

TCHOUYA'A NGOKO ISRAEL CHRISTIAN

Thesis submitted in fulfilment of the requirements for the degree

Master of Technology: Information Technology

in the Faculty of Informatics and Design

at the Cape Peninsula University of Technology

Supervisor: Dr B Kabaso

Co-supervisor: Mr A Mukherjee

District Six Campus

November 2020

DECLARATION

I, **Tchouya'a Ngoko Israel Christian**, declare that the contents of this thesis represent my own unaided work, and that the thesis has not previously been submitted for academic examination towards any qualification. Moreover, it represents my own opinions (apart from normal guidance from my supervisor) and not necessarily those of the Cape Peninsula University of Technology.

Signed **Date**

ABSTRACT

In this new century, the huge amount of data produced daily will remain useless unless we use emerging tools and technologies to make it accessible. There is a need of content summarisers, to reduce manual summarisation which is time consuming and incurs massive costs.

Over the recent years sequence-to-sequence learning has attracted more interest. Text summarisation in natural language processing has been limited to extractive methods that select the important sentences of the original text and combine them to form the final summary. The success of end-to-end training of encoder-decoder neural networks in machine translation tasks has developed research using the same architectures in tasks such as paraphrase generation or abstractive text summarisation.

Abstractive text summarisation attempts to get the main content of a text and compresses it while keeping its meaning, its semantic and grammatical correctness. It generates dynamic paraphrases and produces natural summaries. It has been recently less attempted and understood. These sequence-to-sequence models founded on Recurrent Neural Networks (RNN) were able to link the input and output data in an encoder-decoder architecture. Further producing good output summaries with the inclusion of attention mechanisms to the RNN layers. Research has shown the good performance of these architectures by using attention mechanisms in machine translation. Abstractive text summarisation using recurrent neural networks with attention mechanisms at sentences has produced better results. It has excelled the recent state-of-the-art model of abstractive text summarisation. However, for longer document summaries, these models often contain grammatical errors. In this investigation we employ a data-controlled approach using recurrent neural networks at paragraph level and train the model end-to-end, to predict the summary for a given text document. We evaluate this model to the DUC 2004 datasets. Our model produces higher quality summaries and obtains 44.44 ROUGE-1 score, 22.50 ROUGE-2 score and 45.15 ROUGE-L score on DUC 2004 datasets.

ACKNOWLEDGEMENTS

I would like to thank:

- The Almighty God, may he be glorified through and for this work. Thank you for being the light of my life that never leaves me. I can never thank you enough for your grace and unconditional love.
- My parents, Ngoko Emmanuel and Tchapda Madeleine for everything you have done to make me a better person and for making sure that I get where I am today. Your hard work is never in vain, and I appreciate everything you did and still doing in my life.
- My supervisor, Doctor Kabaso who is more than just a supervisor, he has been my mentor during the last couple of years. Many thanks for easing my understanding of many technical challenges and pushing me to the completion of this work.
- My friends thank you for being who you are in my life and your continuous encouragement.
- My family who encouraged me during these past years.
- The Department of Informatics and Design of the Cape Peninsula University of Technology for providing me with the platform to express my ideas through this research-based Master’s course.

DEDICATION

I would like to dedicate this work to:

- My late father, Ngoko Emmanuel, for being the driving force behind this work and being my inspiration from the beginning to the end of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABS	Abstractive Text Summarisation
AI	Artificial Intelligence
CPUT	Cape Peninsula University of Technology
DUC	Document Understanding Conference
ML	Machine Learning
MT	Machine Translation
NLP	Natural Language Processing
NN	Neural Network
PV-DBOW	Paragraph Vector Distributed Bag of Words
PV-DM	Paragraph Vector Distributed Memory.
RNN	Recurrent Neural Networks
ROUGE	Recall-Oriented Understudy for Gisting Evaluation

CLARIFICATION OF BASIC TERMS AND CONCEPTS

This section clarifies the terms and concepts used in this study.

Term	
Artificial intelligence	Artificial intelligence (AI) coined using two words known as artificial and intelligence (Russell & Norvig, 1995). It is defined as machine intelligence that has the ability of computers to execute tasks that require human intelligence such as speech recognition, translation between languages, etc. (Exner-Stöhr et al., 2017; McCarthy, 2006). Today artificial intelligence has lots of processes in the field such as experts' systems, natural language processing (natural language understanding and generation), robotics, data mining and related fields (Nilsson, 2014).
Natural language generation	Natural language generation (NLG) is a field of AI that can produce meaningful text in languages like English, Spanish, etc. (Gatt and Krahmer, 2018; Reiter & Dale, 1997).
Natural language processing	Natural language processing (NLP) is a part of computer science and artificial intelligence which deal with human languages (English, French…etc.).
Natural language understanding	Natural language understanding (NLU) is mapping the given input natural language into representation (Jhalani & Meena, 2017; Allen, 1995; Schank, 1972).

CHAPTER ONE
INTRODUCTION

Chapter one presents an outline to this thesis. Firstly, it introduces the research area in section 1.1, the background to the research listed in section 1.2, statement of research in section 1.3, hypothesis present in section 1.4, aim and objectives in section 1.5, delineation of research in section 1.6, significance of the study in section 1.7, expected outcomes, results and contributions of the investigation in section 1.8 and in section 1.9 the thesis overview.

1.1 Introduction

Since the release of *World Wide Web*, information has become available in quantity for every topic. In addition, the freedom to publish on the internet has also led to the fast growth of information (Allahyari et al., 2017). Therefore, a demand has arisen among the research community to develop methods that summarises documents (Gambhir & Gupta, 2017). However, automatic text summarisation has pushed researchers to implement methods, which reduce the size of documents while keeping its valuable content and can match human effort (Hou, Hu & Bei, 2017).

Automatic text summary is a collection of information from a source document (Gambhir & Gupta, 2017; Huang et al., 2010; Luhn, 1958). Automatic text summarisation can produce concise and fluent summary (Wang, 2016).

Some of the applications where automatic text summarisation is beneficial include storyline of events, text compression and summarisation of user-generated content (Khan & Salim, 2014). However, automatic text summary is a challenging task. It has concerns such as sentence ordering, redundancy, text cohesion that require consideration when summarising documents (Gambhir & Gupta, 2017; Huang et al., 2010).

Recall-Oriented Understudy for Gisting Evaluation (ROUGE) was presented in 2004 (Lin, 2004). It includes measures which has the potential to automatically evaluate the quality of machine summary compared to a set of human summaries. It has four different measures: a) ROUGE-N which counts the number of n-grams, b) ROUGE-L which counts word occurrences, c) ROUGE-W which weighs the longest common sub-sequence, d) ROUGE-S which skips word-pair in the sentence order (Wang et al., 2016; Lin, 2004).

1.2 Background to the research

The easy access to computerization has led to a massive generation of information. This has given a challenge among numerous users in selecting the important information, often deeply buried in the text that is being produced (Gambhir & Gupta, 2017). Two major techniques of automatic summarisation have been suggested to offer solutions to these challenges, extractive and abstractive summarisation.

Extractive summarisation selects relevant sentences from the source document(s), then combines them to produce a summary. While abstractive summarisation (ABS) interprets the source document(s) and produces a summary (Khan & Salim, 2014). Abstractive text summarisation needs deep understanding of the text to create new phrases, which constitute the generated summary (Dalal, Vipul and Malik, 2013). According to Genest and Lapalme (2012) abstractive summarisation has improved accuracy, reduced redundancy and a good compression rate of a summary.

However, the development of deep learning models in NLP tasks (LeCun et al., 2015), has motivated researchers to use deep neural networks architecture for abstractive text summarisation (Nallapati et al., 2016). Recent Artificial Neural Networks (ANN) models based on the attention encoder-decoder model for machine translation were able to produce abstractive text summaries with better ROUGE benchmark scores (Bahdanau et al., 2014). These systems use sequence-to-sequence, input/output to generate summaries (Paulus, Xiong & Socher, 2017).

Further, according to Rush et al. (2015), RNN model has been successfully used to summarise documents at sentence level. They believed that paragraph level summarisation could improve the quality of the output summary.

1.3 Statement of the research problem

Abstractive text summarisation is an emerging and dominant technique. Nevertheless, it is fraught with challenges known from the ROUGE benchmarking results compared to human summarisation (Nallapati et al., 2016; Chopra, Auli & Rush, 2016; Lin, 2004). Further, the output of the summarisation is often grammatically incorrect (Rush et al., 2015). Since the process uses corpus, the meaning of the output document is a challenging task because it is not always the same with the input text (Moratanch & Chitrakala, 2016). Rush et al. (2015) state that, abstractive summarisation using RNN has produced better ROUGE scores at sentence level. Since sentence level summarisation summarises sentence by sentence, the number of sentences in the summarised document are still conserved. Further, the quality of the summary is still poor. There is a need to reduce the number of summarised sentences and improve the quality of the summary. Research is required to apply RNN based paragraph level abstractive text summarisation to see if this would reduce the number of sentences and improve ROUGE scores.

1.4 Hypothesis

H0. Abstractive text summarisation using RNN at the paragraph level does not produce different number of sentences and does not improve ROUGE scores compared to sentence level abstractive text summarisation.
H1. Abstractive text summarisation using RNN at the paragraph level produces different number of sentences and improves ROUGE scores compared to sentence level abstractive text summarisation.

1.5 Aim and objectives

1.5.1 Aim

- The aim is to improve abstractive text summarisation, using RNN at paragraph level.

1.5.2 Objectives

- To understand the requirements and parameters of abstractive text summarisation at the paragraph level.
- To investigate the application of abstractive text summarisation using RNN at paragraph level.
- To compare the number of sentences and ROUGE scores of abstractive text summarisation using RNN at sentence level with paragraph level.

1.6 Delineation of the study

This research is limited on a single document at the time of each dataset used. It will look at summarise documents at paragraph level, using unsupervised machine learning, delineated to recurrent neural network algorithm.

1.7 Significance of the study

This study is intended to offer an insight sympathetic to the existing body of knowledge with respect to the studies on abstractive text summarisation using recurrent neural networks. It improves the grammaticality and the quality of abstractive text summarisation.

1.8 Expected outcomes, results and contributions of the research

The expected outcomes will be the empirical data and knowledge about the application of RNN at paragraph level. It will contribute to the software code for people to produce good summarised documents. Further this research will help the research community to understand the content using the summarised abstraction in the least amount of time.

1.9 Thesis structure

Chapter one: Introduction

Chapter one gives an overview of problems being investigated in this study. It describes the background of investigation, aim and the objectives, and hypothesis. Lastly, it discusses the delineation, significance, then the result and contribution of the research.

Chapter Two: Literature review

Chapter two presents and explains the content of this research. It provides the insights of abstractive text summarisation using RNN. Then presents the systematic literature review that outlines the review protocol, exclusion and exclusion criteria to determine the selected studies.

Chapter Three: Research methodology

Chapter three discusses the research methodology and design used in this investigation. It includes philosophical assumptions, the research approach and strategy, data collections, data analysis and ethical considerations. It also includes the conceptual framework used in the study.

Chapter Four: Experimental planning, setup and implementation

Chapter four discusses the goals of the experiment and how it was achieved, the participants and materials used. Tasks involved and procedures to execute these tasks, the deviation from the plan to know how the hypothesis will be measured. It also focuses on the experiment. It explains how the data is collected, then test the hypothesis and evaluate the effectiveness of the artifacts being produced in the project. It also gives the deep analysis of the results gathered from the experiment.

Chapter Five: Findings and discussion

Chapter five summarises the research findings by giving the facts obtained from the hypothesis testing process. Furthermore, it discusses the consequent improvements and shortfalls of the system. Finally, it discusses the theories that will help to advance the knowledge in the field of study.

Chapter Six: Conclusion and future direction

Chapter six is the summary of the experiment. It provides all the essential steps taken in the project to produce the evidence elaborated at the end of the research. It then raises issues for further research.

CHAPTER TWO

LITERATURE REVIEW

In chapter one, the research problem has been identified. This chapter presents the Systematic literature review, which forms the background theory for this thesis. Literature review seeks to clarify and evaluate present literature, relevant to the research that has already been conducted (Aveyard, 2014). It draws from academic journals, scholarly articles, websites and published literature sources. This helps readers around the research, to understand the work presented in the later chapters.

Systematic literature review (SLR) is one of the main tools generated to support an evidence-based model. SLR is employed to garner and evaluate knowledge or experiences from different studies related to a specific research question or hypothesis (Okoli, 2017). Systematic literature review was undertaken to:

- Gather and interpret empirical evidence within abstractive text summarisation available.
- Compare the solution with respect to constraints, methods and/or approaches of abstractive summarisation and identify strengths of different solutions found.
- Describe the implications of the findings when giving solutions.

Figure 2.1 below illustrates the research areas of this study. The complete study lies on improving abstractive summarisation of a text using the recurrent neural network algorithm at a paragraph level. The Literature review is structured in a way to explain each area in this research.

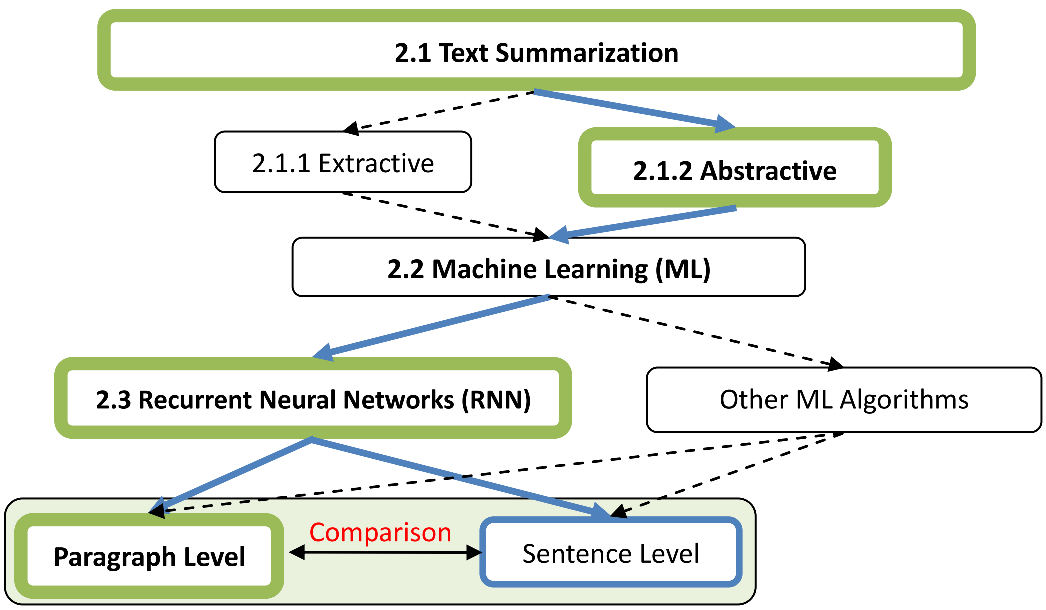


Figure 2.1: Overview of the research

2.1 Text summarisation

The concept of automatic text summarisation started around 1950. Luhn (1958), Gaikwad and Mahender (2016) define it as the process of retrieving or collecting information from a text and presenting the collected information in the form of a summary. According to Barzilay and Elhadad (1999), text summarisation is also defined as: $y_{m} = f(x_n)$ where $x_1, x_{2 \ldots} x_n$ represents the sequence of input (article); where $y_1, y_2 \ldots y_m$ the sequence of output and m < n (where m is the number of input sentence(s) and n is the number of output sentence(s)).

The aim of automatic text summarisation is to reduce the input document in a shorter form that preserves the content and the complete meaning of the original text. This summary or shorter form of text should satisfy the needs of users (Mittal et al, 2014).

Gambhir & Gupta (2017) argued that text summarisation methods are classified in three types, based on the:

- **Input type**: single document and multi-documents.
  - Single document generates the summary of one document input at a time.
  - Multi-documents produce the summary of multiple documents written for the same topic.
- **Purpose type**: generic, domain-specific and query-based.
  - Generic summarisation produces the summary of any input document. The goal is to get the important information from an input document.
  - Domain-specific summarisation generates more accurate summaries.
  - Query-based summarisation generates the summary relevant to the user request. The machine summariser chose from the input information to convey to the output summary.
- **Output type**: extractive and abstractive.

This research is input type as a single document, domain specific as purpose type and abstractive as output type.

At the early stage of research, the researchers focus mainly on extracting keywords from the original text using different algorithms. This process of automatic summarisation was known as extractive text summarisation (Allahyari et al., 2017; Edmundson, 1969; Luhn, 1958).

2.1.1 Extractive summarisation

Extractive summary is done by retrieving the relevant sentences from the document(s), based on statistical analysis such as words/sentences frequency or cue words to extract sentences

(Kyoomarsi et al, 2008). This summarisation technique is easy to implement because not a deep understanding of text is required. Extractive summarisation has been proven less coherent and not effective for the generated summarised text (Sunitha, Jaya & Ganesh, 2016). Recently researchers started to study abstractive summarisation using various algorithms (Huang et al., 2010; Goldstein, 2000).

2.1.2 Abstractive summarisation

Abstractive text summarisation is a task of reducing a document to a version shorter than the original document(s) (See et al., 2017). It is aiming to interpret and examine the original document(s) and generates a concise summary (Hou et al., 2017; Dalal, Vipul and Malik, 2013). Abstractive text summarisation helps to reduce the work needed to digest large volumes of text (reduce reading time, ease document selection and content in small devices). It generated good summaries, which content tool that enhances the original content (Condori, López & Thiago, 2017; Salim, 2014). It has the following benefits.

- It gives an intelligent, consistent and less redundant summary.
- It uses both natural language processing (NLP) techniques (semantic representation and NLG) and compression techniques (Jhalani & Meena, 2017; Dalal, Vipul & Malik, 2013).

Abstractive text summarisation is grouped in two categories: structure-based and semantic based approaches.

2.1.2.1 Structure-based approach

It encodes essential information from document(s) through intelligent schema such as scripts and frames (Kasture et al., 2014). This approach comprises of the rules-based method, tree-based approach, ontology-based method, lead and body phased method, template-based method and graph-based method (Moratanch & Chitrakala, 2016).

2.1.2.1.1 Tree-based method

This method uses a dependency tree to represent the text and an algorithm for generation of the summary. However, it needs a model which includes a representation for content selection (Hirao et al., 2015).

2.1.2.1.2 Template-based method

Template based method uses a template to represent an entire document text. Linguistic patterns or extraction rules correspond to identify text snippets mapping into template slots. The generation of summaries is highly coherent since it relies on relevant information by the information extraction system. However, it needs to design templates and generalize these templates which is difficult (Oya, 2014).

2.1.2.1.3 Ontology-based method

The ontology-based method improves the summarisation process using knowledge-base. It also uses fuzzy ontology to manage unknown data that simple domain ontology cannot. The ontology-based method eases to draw relation or context. However, it is only used to Chinese news and generating the rule-based system to handle uncertainty (Ragunath & Sivaranjani, 2015; Ramezani et al., 2015).

2.1.2.1.4 Lead and body phrase method

Lead and body phrase method is grounded on insertion and substitution of sentences having the same syntactic head chunk in the lead and body sentences. This allows to rephrase the lead sentence. It is good for semantically revising a lead sentence. However, analyzing errors degrade the completeness of a sentence such as grammar and repetition. Lead and body phrase focus on rephrasing techniques and lacks an entire model which would contain an abstract representation for content selection (Gaikwad & Mahender, 2016).

2.1.2.1.5 Rules-based method

In the rules-based method, articles are represented in categories and list aspects to be summarised. This method generates summaries with better information density. The limitation is that rules and patterns are tedious and time consuming because they are manually written (Genest & Lapalme, 2012).

2.1.2.2 Semantic-based approach

This approach uses semantic representation of document(s) for NLG systems (Greenbacker, 2011). It determines verb and noun phrases when processing linguistic data. Then link phrases to attributes and relations of specific ontology. Further, important sentences are identified using clustering techniques and ontology-based annotation. At last, final information are used to convert into semantic representation which feed to NLG system and produce abstracts (Elsied and Salim, 2013). According to Khan and Salim (2014), structure-based approach produces short, coherent, information rich sentences with little redundancy. While semantic based approach does the same and improves the linguistic quality of the summary. Semantic based will be used in this research for its advantages.

The semantic-based approach comprises the multimodal semantic model, information item-based method and semantic graph model (Moratanch & Chitrakala, 2016).

2.1.2.2.1 Multimodal semantic model

Multimodal semantic model uses a semantic model to capture concepts and relationships then represent the contents of multimodal documents. This framework produces abstract summaries with an excellent coverage because it includes graphical content and salient textual of the entire document, but it is manually evaluated by humans (Greenbacker, 2011).

2.1.2.2.2 Information item-based method

The summary is produced from abstract representation of document(s) not from sentences of document(s). This method produces a coherent, information rich, short and less redundant summary. However, it is difficult sometimes to create meaningful and grammatical sentences from abstract representation. The linguistic quality of summaries is not good due to incorrect parses (Genest & Lapalme, 2012; Mallett, Elding & Nascimento, 2004).

2.1.2.2.3 Semantic graph model

Semantic graph model uses semantic representation to feed into NLP. Its aim is to summarise document(s) by creating a rich semantic graph for the source document, reducing the generated graph, then generating the abstractive summary from the reduced semantic graph. In addition, it is concise, coherent, reduces redundancy and offers grammatical correctness of sentences. Semantic graph model summarisation can achieve fifty percent performance compared to human summarisation (Moawad & Aref, 2012). It is the best semantic-based approach; however, it still needs grammatical improvement. Semantic graph is limited to single document summarisation.

2.2 Machine learning

Machine learning (ML) is the design of algorithms that allow computers to learn and perform tasks, using intelligent software (Ayodele, 2010). It is a subfield of AI, lying at the intersection of statistics and computer science, the score of data science and AI (Jordan & Mitchell, 2015). Shalev-Shwartz and Ben-David (2014) consider ML as a science which focuses on algorithm models that allow computers to learn without being explicitly programmed. ML has the potential to become the mainstream of business because it is used to reduce cost, increase profit, improve customer experience, and save lives (Rudovic et al, 2018; Brynjolfsson & Mitchell, 2017). ML is a common tool in tasks that are complex to programs requiring information extraction in large data sets (LeCun, Bengio & Hinton, 2015). It conducts predictions, performs clustering and makes decisions from a given dataset at the same level as humans. According to Schuld, Sinayskiy and Petruccione (2015) ML plays a major role in real world applications such as:

- Computational finance for credit scoring.
- Image processing and computer vision for face recognition and object detection.
- Energy production for forecasting and price.
- Automotive and manufacturing for predictive maintenance.

Nevertheless, ML requires human intervention because it cannot do the entire task human beings can do. It is categorized into different techniques which are supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning (Mohammed, Khan & Bashier, 2016).

2.2.1 Supervised learning

Supervised learning is a machine task of deducing a function from a trained data provided by a human being. Its purpose is to get a computer to learn classification systems that have been created and produce the correct output (class label or real number) from an input (Ayodele, 2010). It is a technique for training neural networks and decision trees. According to Donalek (2011), in supervised learning:

- Training data include input and desired results.
- The correct output from a given input during the learning process.
- It is fast and accurate.

Supervised learning is grouped in two algorithms: regression and classification.

- ➢ Classification techniques predict the discrete responses like, to know if an email is genuine or spam, then inputs data into categories in applications. Examples of applications: medical images, speech recognition and credit scoring. Classification algorithms are support vector machines, naïve Bayes …etc. (Tantithamthavorn et al., 2016).
- ➢ Regression techniques predict continuous responses such as change in temperature and fluctuations in power demand. Examples of applications: load forecasting and algorithm trading. Examples of regression algorithms are decision tree, linear regression, neural networks, ensemble methods (Gong, 2018).

2.2.2 Semi-supervised learning

Semi-supervised learning is the combination of classified (labelled) and unclassified (unlabeled) data used to produce a model for the classification of data. Its goal is to study the model predicting classes of upcoming test data better than a model using only labelled data (Sutskever et al., 2015). It often improves the performance of supervised learning tasks for which there is not large data.

Nevertheless, it is difficult for unsupervised cost function to improve the performance of supervised cost function, due to the lack of labelled data.

2.2.3 Reinforcement learning

Reinforcement machine learning aims to use observation and learns through trial and error(s) interactions of the environment, then actions to minimize the risk and maximize the future rewards received over the lifetime. In this learning approach problems studied are formally equivalent and solutions are the same, despite different aspects of the problem. It is used mainly for decision and control field theory (Zoubin, 2004).

2.2.4 Unsupervised learning

Unsupervised machine learning goal is to model the distribution or underlying structure in data to learn about data. In other words, have the computer do something that no one tells it to do. It identifies hidden patterns or intrinsic structures in input data (unclassified or unlabeled data) provided by the environment. Unsupervised learning does not require any training data (labelled data), the machine can generate output by accessing the original input (Gambhir & Gupta, 2017; Donalek, 2011). This technique is done by the teaching agent using the rewards system to indicate success (classify data), or by finding similarities in the training data or cluster data.

Some of the benefits of unsupervised learning algorithms which motivated this research are the following:

- Unsupervised learning is learning without an instructor.
- Cluster the input data in classes and provide precise results during training (Donalek, 2011).
- The algorithm will try to identify a pattern in the inputs and categorize inputs that have a similar pattern (Marsland, 2015).

However, it is difficult to measure the accuracy of unsupervised learning model and there is no point where you need to stop when using hierarchical algorithm.

Samples of unsupervised machine learning are self-organizing maps, independent component analysis and multi-dimensional scaling (Lu & Wang, 2010).

In summary, reinforcement machine learning assumes that an agent operates in an environment and receives a reward after performing an action many times. Semi-supervised and supervised machine learning approaches are used when partial or complete truth is available, such as label for classification problems and real value for regression problems. While unsupervised learning is used when there is no truth, its goal is to identify structures in the input space used to decompose the problem and facilitate the model building (Bonissone, 2015).

This research uses unlabeled data. Hence, the research is focused on using unsupervised machine learning algorithm.

Popular algorithms in unsupervised learning are K-means, hierarchical clustering, neural network, Gaussian mixture and hidden Markov model learning algorithm (Celebi, Emre & Aydin, 2016).

2.2.4.1 Hierarchical clustering

This algorithm groups similar items into clusters (groups). It finds successive clusters using previous clusters established. It can be, either agglomerative (bottom-up) or divisive (top-down) algorithm (Rokach, Lior & Maimon, 2005).

- In agglomerative algorithm, each element begins at separate clusters and merges successively to form a large cluster. This produces grouping at the next higher level with one less cluster. This algorithm is generally faster to compute. It is used in market segmentation, image processing, recommendations systems and search engine results analysis. Figure 2.2 illustrates agglomerative algorithm.

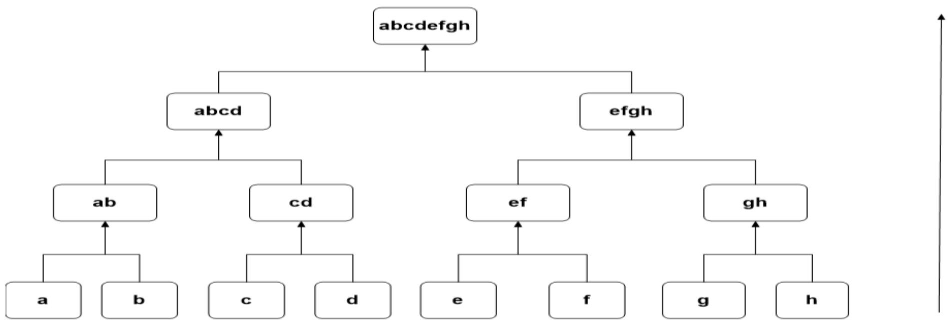


Figure 2.2: Agglomerative algorithm (Erman, Korosec & Suklan, 2015)

- Divisive algorithm determines all clusters at once. It is less ‘’blind’’ to the global structure of the data.

2.2.4.2 Hidden Markov Model

Hidden Markov Model is a model used to represent probability distributions over sequences of observations. It is most important for unsupervised machine learning models in speech recognition systems and data compression (Ghahramani, 2001). It allows us to speak about observed events (input words) and hidden events (parts-of-speech tag) that we think of as casual factors in our probabilistic model. This model is specially used in reinforcement machine learning algorithms. Figure 2.3 illustrates a sample of Hidden Markov Model.

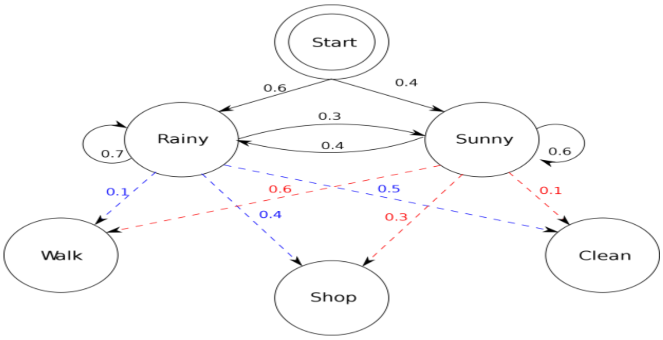


Figure 2.3: Hidden Markov Model (Kang, 2017)

2.2.4.3 K-means

K-means is an unsupervised technique, the most popular and powerful clustering algorithm due to its simplicity and efficiency. It performs better than other algorithms. It is used for problems in which all variables are the quantitative type and square Euclidean distance (Sutskever et al., 2014). K-means algorithm is used for clustering large datasets to classify objects into k numbers of groups. Nevertheless, it is only applicable when the means is defined. The value of the number of desired clusters is essential to be given as an input, irrespective of the distribution of the data points (Nazeer, Abdul & Sebastian, 2009). This algorithm is not used for prediction, but to group data.

2.2.4.4 Gaussian mixture

Gaussian mixture model is the statistical model for grouping data with real-value components. It shares key similarities with k-means but aims to maximize the probability function of the parameters comprising the means, covariance and the mixing coefficient (Sridharan, 2014).

2.2.4.5 Neural networks

Ideally, neural networks refer to the biological human brain; the computational model is Artificial Neural Networks (ANN). It is the mathematical model designed to work the way the human brain analyses and processes information (Gurney, 2014, p.4). It is a network diagram which has one input layer, one or more hidden layers and one output layer. It can perform tasks such as classification, decision-making, prediction, visualization etc. ANNs are powerful learning models that realize in unsupervised and supervised learning tasks the state-of-art (Lipton et al., 2015). Figure 2.4 shows a sample of ANN.

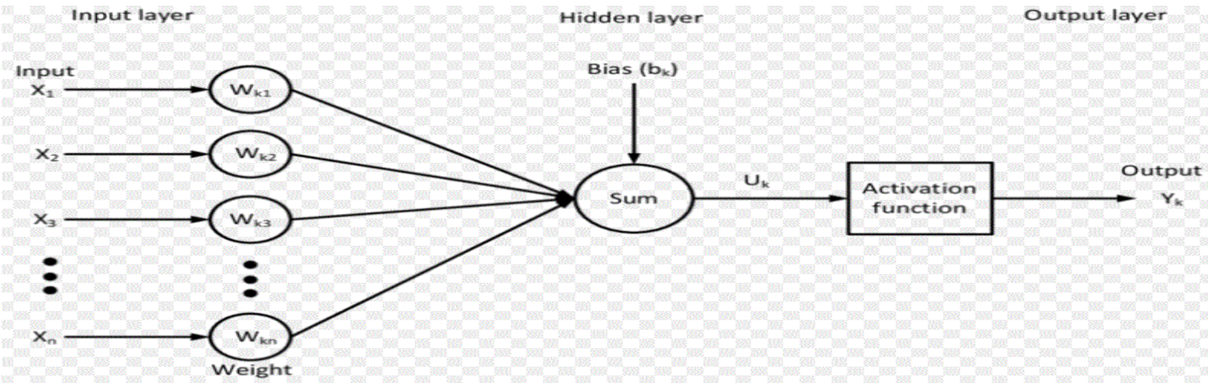


Figure 2.4: Artificial neural network (ThippeSwamy & RashmiAmardeep, 2017)

There are varieties of ANNs. The most used are feedforward and recurrent neural network (Krisel, 2007).

2.2.4.6 Feedforward Neural Network

Feedforward Neural Network (FNN) is an ANN in which links between nodes do not form a cycle. Each neuron in one layer has only direct connections to the neurons of the next layer (Krisel, 2007; Zell, 1994). The information moves from the input nodes to the output nodes. Figure 2.5 shows a sample of a FNN.

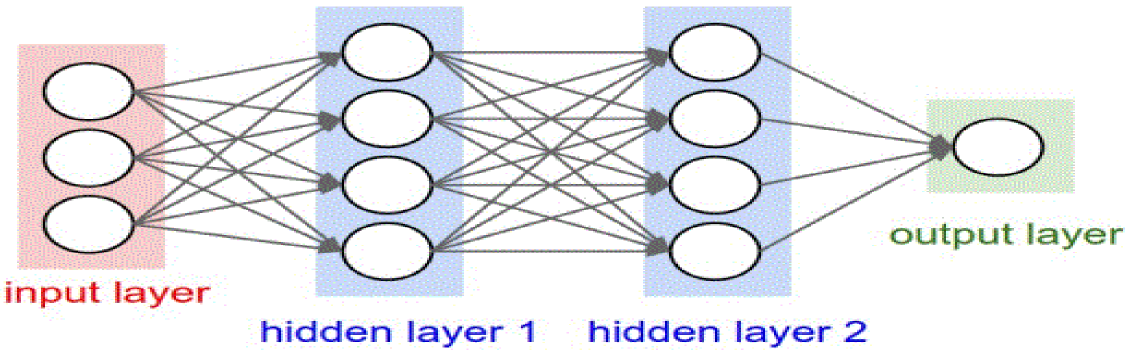


Figure 2.5: Feedforward neural network (Haytam, 2019)

2.3 Recurrent Neural Network

Recurrent Neural Network is an ANN having connections between nodes forming a direct graph. It is adapted from a standard feedforward neural network (Sutskever et al., 2014; Sundermeyer et al., 2012). Recurrent Neural Networks (RNN) are models producing excellent results on many tasks such as summarisation, data modelling and statistical analysis. Figure 2.6 illustrates a sample RNN.

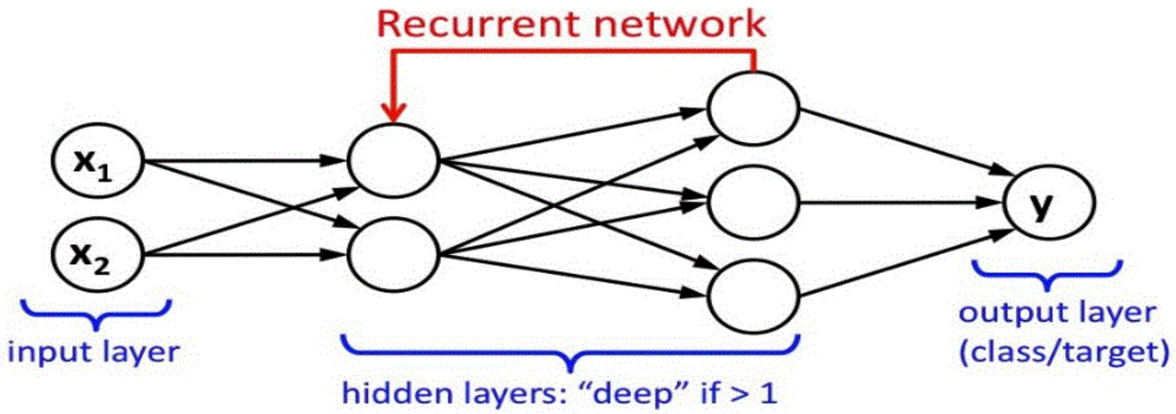


Figure 2.6: Recurrent neural network (Bullinaria, 2013)

In summarisation, RNN scores sentences are fairly based on probability of occurrences in the input paragraphs. It gives a measure of grammatical and semantic correctness of a sentence (Logeswaran, Lee & Radev, 2016). It performs calculations on sequential data. RNN has become a standard for NLP tasks. Researchers often use early stopping, small and under-specified models, because the large RNN tends to overfit (Zaremba et al., 2014).

RNN can learn to reduce the full history in low dimensional space, while FNN compresses only a single word. In addition, RNN can form short-term memory to deal better with position invariable, but FNN cannot do it (Mikolov et al., 2010). According to Sundermeyer et al. (2012), RNN allows to improve an application in which it is used.

RNN use many encoder-decoder models for articles compression, for sequence prediction problems that have a variable number of inputs, outputs or both. The most used are attention-based, convolutional and bag-of-words encoder-decoder models (Rush et al., 2015).
Bag-of-words model capture the value of words to differentiate content words from stop words. It learns to combine words and it is limited to represent joining phrases (Bahdanau et al., 2014).

Convolutional encoder-decoder model improves bag-of-words by supporting local connections between words; encode the input sentence while not requiring the context (Rush et al., 2015).

Attention-based model is the model that allows the machine to look through information in the original sentence, holds it, then generates the word according to the current word it works on and the context. It solves the translation challenge that relies on reading a complete sentence and compresses information into a specific length vector (Cho et al., 2015; Bahdanau et al., 2014). A sentence with hundreds of words will lead to inadequate translation or information loss.

Rush, Chopra and Weston (2015) argue that, convolutional encoder has better capacity compared to bag-of-words. However, it needs to output single representation for an input sentence. Bahdanau et al. (2014) states that, machine translation uses attention-based encoder to create a representation.

For the above-mentioned advantages of the attention-based model, this research focuses on using the RNN with encoder-decoder model as the tool for generating abstractive text summarisation of single documents.

2.4 Summary of research literature

A literature review on abstractive text summarisation has found that there are numerous researches done on abstractive text summarisation, using recurrent neural network (See, Liu & Manning, 2017; Chopra, Auli & Rush, 2016). These have produced an efficient and good compression ratio for the generated summary.

Recent researches has proved that with the used of machine learning algorithm such as recurrent neural network with attentional encoder-decoder, abstractive text summarisation produces excellent results (Sutskever et al, 2014).

Rush et al. (2015) state that, abstractive text summarisation using recurrent neural network has produced better ROUGE scores at sentence level. However, the number of sentences of input text and output text is the same and the ROUGE scores is still far below the performance of human summarisation. Research is requiring being on abstractive text summarisation using recurrent neural network at paragraph level.

2.5 Systematic literature review of abstractive text summarisation using recurrent neural networks

Text summarisation started in 1950. To ease the challenge of summarisation, most techniques follow extractive summarisation, by generating the important sentences of a document without any change (Abbas, Elsied & Salim, 2013). Although it is a well-known method, the final summary presents many problems, such as lack of coherence and consistency. The abstractive summarisation was introduced to solve these limitations since it produces a summary from a fragment of information. In recent years, research in ABS has been criticized because of information being repeated in the generated summary and in the way that present information is required. Sentence compression, sentence fusion and natural language processing has applied to generate an abstract (Radev and McKeown 1998; Barzilay & McKeown 2005). A study in terms of coherence of summaries generated, show that 78% of abstractive summary were more coherent than extractive summary (Lloret & Palomar, 2012).

There has been improvement in ABS techniques for the previous years. This reviews contents of the previous state-of-the-art. It aims to enhance the understanding of the literature on ABS, by determining techniques and performance; identifying research gaps; and offer recommendations.

2.5.1 Research questions

Research question(s) must be clearly specified in the systematic literature review to give a clear research scope. The SLR goal is to collect and investigate solutions that have been used in abstractive summarisation using recurrent neural network. The research questions (RQs) are:

RQ1: What are the existing methods previously used for abstractive text summarisation?

RQ2: How are abstractive text summarisation methods benchmarked?

RQ3: What are the drawbacks of the existing methods used?

2.5.2 Review protocol

Review protocol helps the researcher to avoid prejudice that could negatively influence the objectives and goals of the systematic review process. Okoli and Schabram (2010) argue that it is an important step to the SLR. It constitutes search strategy, study selection, quality assessment, data collection and data analysis.

2.5.2.1 Search strategy

Search strategy is used to formalize the list of possible sources (digital libraries) that could provide relevant literature to the systematic review. The table below enumerates selected digital libraries used to identify keywords and/or search terms.

Table 2.1: Digital libraries for search strategy

Source	URL	Responsible
ACM Digital Library	https://dl.acm.org/	Israel
IEEE Xplore Digital Library	http://ieeexplore.ieee.org/Xplore/home.jsp	Israel
SpringerLink	https://link.springer.com/	Israel
Scopus	https://www.scopus.com/	Israel

The search string(s) is carried out by selecting the most relevant keywords based on the research questions and the topic investigation. We also include synonyms and alternative words in the search string. The search was made from 2010 till 2019 to retrieve the most recent papers of abstractive text summarisation. Digital libraries offer the possibilities where keywords and search string(s) can be entered. They also offer advanced search options for users to formulate search strings with conjunctions (AND and/or OR).

The search is based on the title, abstract and keywords. Each category offers different research literature. A combination of all the categories provide the search string that finds the relevant studies needed for this systematic literature review.

Table 2.2: Search terms

	Category 1	**Category 2**
Phrase 1	Abstractive summarisation	Neural networks
Phrase 2	Abstractive text summarisation	Recurrent neural networks
Phrase 3	Text summarisation	

2.5.2.2 Database used

The list of databases used is the following: ACM Digital Library, IEEE Xplore Digital Library, SpringerLink and Scopus.

2.5.2.3 Study selection

Study selection has aimed to select relevant studies related to the research. The following are the inclusion and exclusion criteria of this SLR.

2.5.2.3.1 Inclusion criteria:

- Empirical studies on the abstractive summarisation using recurrent neural networks.

2.5.2.3.2. Exclusion criteria

- Studies with duplicate title or that are not relevant to the review.
- Studies before 2010 or not in English.
- Papers not backing up with evidence or experiments to support the claim.
- Extractive text summarisation.
- Studies either than journal or conference papers.
- Review not relevant to the studies.

After the first step of filtering only 64 studies remained. A set of inclusion/exclusion (keywords and/or full text evaluation) and quality screening criteria were developed to filter out irrelevant studies.

The study is on abstractive text summarisation using RNN. Consequently, the terms such as "abstractive text summarisation" OR "abstractive summarisation" AND "neural networks" OR "recurrent neural networks" must be specified in the keywords, abstract or title. It allows for collecting the first results of the research terms. After carefully screening and selecting papers from the first filtering, studies that are relevant to the full text research, 57 papers were rejected and only 7 papers passed onto the full inclusion and exclusion criteria. Papers selected were given a unique number **(P)** to facilitate their representation in the document as they are sorted in descendant order of publication year in table 2.4 below.

Table 2.3 describes the outcome of the screening process at each stage of execution of the search string within each digital library.

Table 2.3: Search string results

Sources	Number of studies	Studies relevant to keywords of the title	Studies relevant to the full-text research
ACM Digital Library	27	25	1
IEEE Xplore Digital Library	38	7	1
SpringerLink	110	8	1
Scopus	181	24	4

2.5.3 Study quality assessment

Each systematic literature review should have a quality assessment to evaluate the quality of papers selected, their importance and rate for each research question. In the review process the quality assessment steps include the following: the research statement of aim; the understanding of the study if it is associated with other related research and the research method must be described. Table 2.5 shows the ten-quality assessment (QA) questions used in this review process which were adopted from Malhotra (2015). Scores are given by assigning 1 for strongly agree, 0.5 for partly, and 0 for disagree. Studies with scores less than 7.0 out of 10 for QA questions were rejected.

2.5.4 Data extraction

The data collection process was developed by the researcher to ensure that all information relevant to the research questions is extract from the final papers selected in the primary studies.

Table 2.4: Papers selected from primary study

Paper number	Author's names	Publish Year	Paper's title
P1	See, Abigail, Peter J. Liu, and Christopher D. Manning	2017	Get to the point: Summarisation with pointer-generator networks.

P2	Zhou, Qingyu, Nan Yang, Furu Wei, and Ming Zhou	2017	Selective encoding for abstractive sentence summarisation
P3	Hou, Liwei, Po Hu, and Chao Bei.	2017	Abstractive document summarisation via neural model with joint attention
P4	Chopra, Sumit, Michael Auli, and Alexander M. Rush	2016	Abstractive sentence summarisation with attentive recurrent neural networks
P5	Rossiello, Gaetano, Pierpaolo Basile, Giovanni Semeraro, Marco Di Ciano, and Gaetano Grasso	2016	Improving neural abstractive text summarisation with prior knowledge
P6	Alexander M. Rush, Sumit Chopra and Jason Weston	2015	A neural attention model for abstractive sentence summarisation
P7	Moawad, Ibrahim F., and Mostafa Aref	2012	Semantic graph reduction approach for abstractive text summarisation

2.5.5 Results

We present the synthesis of the evidence of our systematic review with the literature search results. During the process of selection titles and abstracts, conclusions and recommendations were screened for important check of the sources. Full papers were taken whenever the minimum requirement for the inclusion criteria was met which was otherwise excluded for the study. Table 2.4 shows the results of papers selected, in order by year of publication.

2.5.5.1 Quality evaluation

The quality evaluation will be moderate according to the quality assessment option where yes = 1, no = 0 and partially = 0.5. The table below gives the result of each paper assessed.

Table 2.5: Quality evaluation of selected papers

Study	Research statement of aim	Study understood if it is associated with other related research	Research experiment is described	Total
P1	1	1	1	3
P2	0.5	1	1	3
P3	1	1	1	2.5
P4	1	1	1	3
P5	1	1	1	3
P6	1	1	0	2
P7	1	1	0	2
Total	6.5	7	5	18.5

2.5.5.2 Abstractive text summarisation

The following section presents a summary of different abstractive text summarisation from the set of papers included in this review.

In **P1** Abigail, Liu and Manning (2017) present an architecture to improve the attentional model, using the hybrid pointer-generator and coverage to keep record of sentences which have been summarised to discourage repetition. Pointer network model uses attention distribution of Bahdanau et al. (2015) to generate the sequence of output from an input sequence. According to Vinyals et al. (2015), hybrid pointer-generator network facilitates the copy of words from original document via pointing, improves accuracy and handling of out-of-vocabulary words, although retaining the capacity to generate new words. This helps to decrease the replication of inaccurate details, avoid repetition and surpass the best abstractive model (Abigail, Liu and Manning, 2017). Yet, the level of abstraction still remains an issue.

Zhou et al. (2017) in **P2** built a selective encoding model to extend the sequence-to-sequence framework for abstractive sentence summarisation. This encoder consists of sentence encoder, sentence decoder both built with RNN and a selective gate network. The gate network is then used to select the encoded information and construct the second level of representation, while the sentence encoder reads the input words through an RNN to construct the first sentence representation. At the end, the attention-equipped decoder will generate the final summary from the second level of sentence representation.

Hou et al. (2017) proposes in **P3** an attention mechanism to avoid undesirable faults, repetitive contents, inability to handle out-of-vocabulary words correctly and use the sub-word method to deal with unknown words. The joint attention will be used to store and complete information from an input of each decoder time step (Bahdanau et al., 2014), while avoiding repeated phrases in the output sequence by reviewing previous output information. The joint attention mechanism improves the performance of traditional models and achieves the best performance in single document summarisation. This model is still not working well with multiple documents.

Chopra et al. (2016) built a model to generate the summary of an input sentence in **P4**. The model is an extension of Rush et al. (2015) and comprises a conditional RNN which takes a conditional input from the output encoder module. The encoder of this model is more sophisticated because it explicitly encodes the position of information of each input word. This novel convolutional attention based conditional RNN model improves the performance of abstractive sentence summarisation. However, summarising a text remains an open issue for this model.

In **P5** Rossiello et al. (2016) proposes an improvement of neural ABS using prior knowledge. This model learns the representation of relationships of words in the input document and the output summary, using simple handcrafted linguistic features. It solves the following limitations of neural attention models for abstractive summarisation:

- Large amount of training data to get clear representation that maps good alignment between original text and the related summary.
- Availability to train the models in specific domains.

This model reduces the amount of data in training phase, avoids grammatical errors and produces better summaries. Nevertheless, the generation of ABS from multi-documents remains a challenge.

Rush, Chopra and Weston (2015) proposes in **P6** a fully data-managed method for abstractive sentence summarisation. It uses a local attention-base responsible for the creation of each word of the summary trained on the input sentence. This approach includes linguistic structure but easily scaled to train a huge amount of data. This probabilistic model combines with a generation algorithm to produce accurate summaries. The grammaticality of summaries and the generation of summaries at paragraph level remain a challenge.

Moawad and Aref (2012) present in **P7** a method to generate an abstractive summary for a single article using a rich semantic graph (RSG) reducing approach. Abstractive summarisations require the advanced language generation technique to produce a generalized summary (Das & Martins, 2007). RSG is ontology-based representation used as an intermediate representation of NLP applications. RSG consists of three phases: Creation of RSG to represent the input of the original document semantically, reduce the generated RSG to a more abstract graph and generate the abstractive summary from reduced RSG. Semantic graph reduction approach uses for ABS minimizes the original text to fifty percent.

2.5.6 Analysis

2.5.6.1 **RQ1**: What are the existing methods previously used for abstractive text summarisation?

According to the authors of these papers they are mainly classified into two abstractive text summarisation groups: an approach using prior knowledge or structure-based approach and an approach using NLP generation or semantic-based approach.

A. Structure-based approach encodes important document(s) through psychological feature schema and alternative structures. It comprises the following approaches: rules-based method, tree-based approach, ontology-based method, lead and body phased method, template-based method and graph-based method.

B. Semantic-based approach uses a linguistic diagram of article(s) to feed into NLG system. There are four semantic-based approaches: multimodal semantic model, information item-based method, semantic graph model and semantic text representation model.

2.5.6.2 **RQ2**: How is abstractive summarisation methods benchmarked?

Table 2.6 shows the ROUGE benchmarking scores and the different datasets used in abstractive text summarisation from selected papers.

Table 2.6: Comparative study ROUGE scores of selected papers

Papers	Dataset(s)	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L
P1	CNN / Daily Mail	39.53	17.28	36.38		
P2	Gigaword	46.86	24.58	43.53		
	DUC 2004	29.21	9.56	25.51		
	MSR	25.75	10.63	22.90		
P3	NLPCC2017	34.94	21.17	14.49	11.28	30.66
P4	Gigaword	33.10	14.45	30.71		
	DUC-2004	28.55	8.79	24.73		
P6	Gigaword	30.88	12.22	27.77		
	DUC-2004	26.55	7.06	22.05		

2.5.6.3 **RQ3**: What are the drawbacks of the existing methods used?

Selected papers proved that, abstractive summarisation has not yet found a suitable method that will completely handle redundancy and coherency of a document with maximal ratio compression. The greatest challenge of abstractive summarisation is that the ability of the system is limited by the richness of their representation. Abstractive summarisation methods have shown major improvement in many aspects. They can produce highly cohesive, less redundant and coherent summaries. However, more research still needs to be done.

2.5.7 Summary

A deep study has been done on ABS to find the possible solutions. This allows a comprehensive search over multiple sources that can be reproducible by other researchers. After outlining the research terms and strategies that include search engines, search dates and outlining all inclusions and exclusions, seven papers have been identified for ABS which discussed RNN. A comparison was done to align the research questions with the respective objectives, performance and target audience of abstractive summarisation technology. These papers used diverse algorithms for the implementation of abstractive summarisation.

Conclusions in **Table 2.6** show that the best summarisation progress is still at 46 percent compared to the human summarisation performance.

2.5.8 Limitations and recommendations

Section 2.5 presents a systematic literature review which is targeted at studies of abstractive methods (techniques) to generate ABS. According to P7, graph technique is the most used and dominant summarisation. It proved to be more efficient and guarantees a high ratio of coherence for the summary generated. P7 achieved fifty percent accuracy compared to human summarisation. It is therefore recommended to focus on graph technique for document summarisation. Further, more effort must be made to this approach in the way of picking relevant keywords and the use of good parsers.

2.6 Chapter summary

This chapter presents two major techniques of text summarisation such as extractive and abstractive summarisation. Abstractive summarisation compared to extractive has improved accuracy, reduced redundancy and good compression rate of summaries (Genest & Lapalme, 2012). A systematic literature review on abstractive text summarisation is been discussed. It shown that abstractive text summarisation using RNN with attention encoder-decoder at sentence level produce good results, but still need to be improve the grammatical quality of summary and ROUGE scores since the machine summarisation is still far below human summarisation.

.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

This chapter's purpose is to outline the research design and the methodological approach employed in the research study. The research design focuses on the planned research study specific outcomes. The process and methodology to follow is therefore chosen to support outcomes and the result's significance.

Section 3.2 discussed the conceptual framework followed in the study. Section 3.3 provides the research design, which includes the philosophical position adopted by the researcher, the research approach, methodology and research strategy employed in this study; the data collection and analysis procedures. Section 3.4 outlines the ethical considerations of the study.

3.2 Conceptual framework

A research theory tries to understand and/or explain phenomena. Conceptual framework, as opposed to theoretical framework, describes the concepts and relationships used in this research. It offers a strong base for a researcher to investigate the given phenomena (Cavana et al., 2001). It includes the key aspects of the study field. It is also used to guide the procedures of data collection and facilitate the interpretation of the findings (Smyth et al., 2004).

In this study, the researcher investigates the characteristics and factors of abstractive text summarisation. The summarisation process of information using the unsupervised learning model. Our proposed conceptual framework used RNN algorithm with attention encoder-decoder model to encode and decode articles from the dataset. The RNN algorithm starts with an existing model that is driven by unsupervised machine learning. It trains a machine to analyse articles, then generate the final summary and ROUGE scores at paragraph level. The RNN algorithm continues being adjusted in the training process until we reach the level of satisfaction from the experiment. An open source corpus will be identified for this experiment and the Python programming language will be used to develop this application.

Moreover, outside the evaluation of the model, figure 3.1 has the input and output block with the expected results, namely the ROUGE scores and the summarised articles (including number of sentences) of abstractive text summarisation at paragraph level. This allows the researcher to evaluate and understand the gaps between abstractive text summarisation at sentence level and the paragraph level.

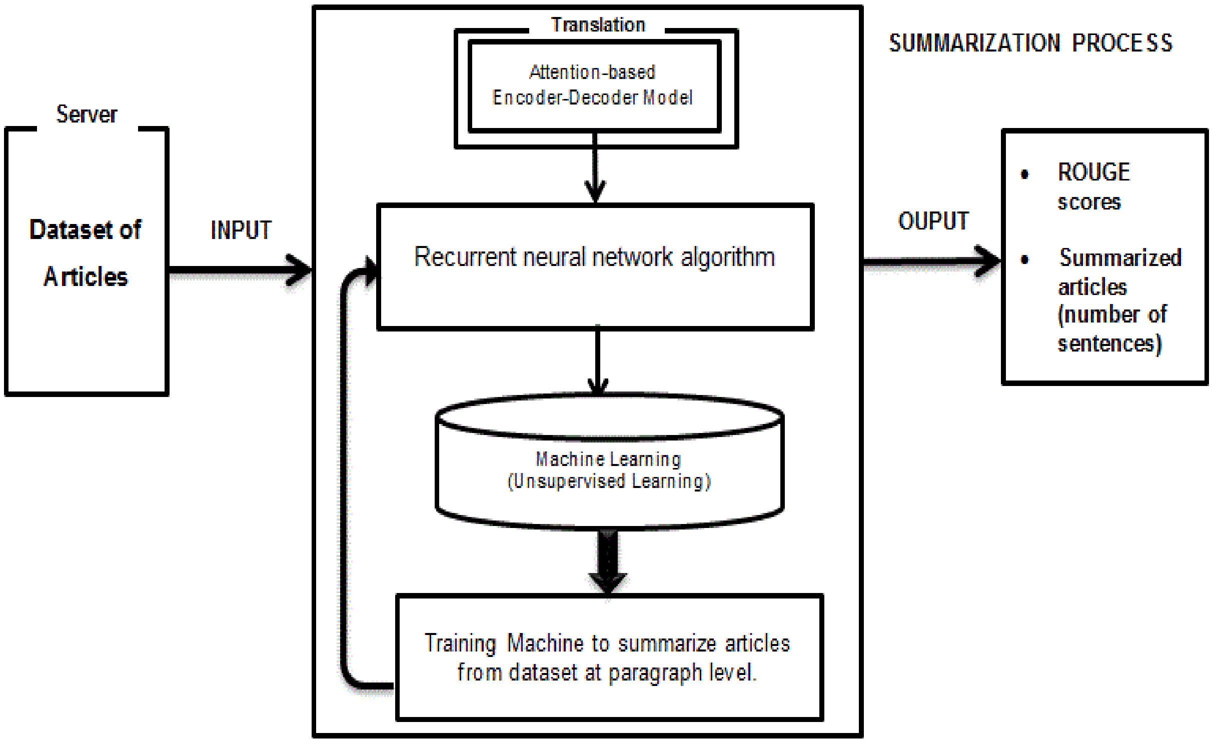


Figure 3.1: Conceptual framework

3.3 Research design

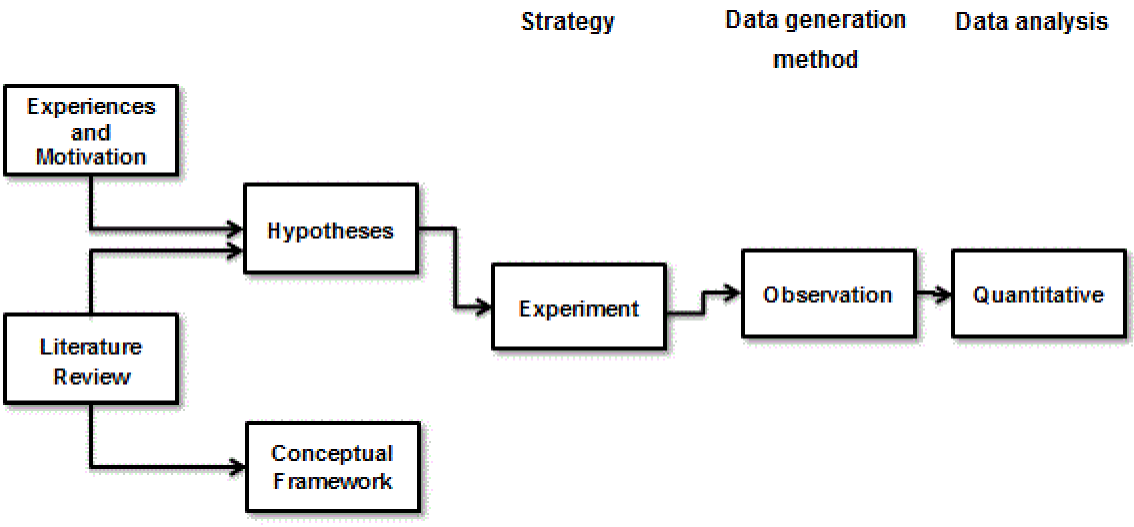


Figure 3.2: Model of the research process (Oates, 2005, p.33)

Research design is considered as a blueprint of a research study. It clarifies how the parts of the research process are connected and organized in a way to lead answers of the questions or hypothesis of the research. Research design should contain research methods and approaches, also suitable procedures and strategies to experiment data collection and data analysis (Bhattacherjee, 2012; Warden, 2011).

The choice of research design is influenced by the research problem, research questions or hypothesis, the researcher's personal experiences, the philosophical assumptions of the investigator and by the audience of research (Saunders et al., 2009). Research is designed according to the aim for which it is proposed to achieve. It depends on whether the aim is to describe (descriptive research), explore (exploratory research) and explain (explanatory research) a phenomenon (Neuman, 2013).

Exploratory research is used when a researcher attempts to observe and understand a new subject of interest or to examine persistent phenomenon (Babbie, 2011). According to Mlitwa (2011), exploratory research is used when investigated phenomena with little or no information is known. Babbie (2011) also states that, exploratory research is conducted when you want to have a good understanding of the phenomena, test the feasibility of widespread research, and to develop quantitative frameworks to be used in conducting future studies. It addresses the ''what'' question revealed Neuman (2013), which makes it the foundation of conducting appropriated scientific research.

Descriptive research is similar to exploratory research. Nevertheless, it helps to describe a phenomenon or topic in detail. It aims to describe the events and situations that cause an effect or outcomes by making observations (Babbie, 2011). It is mainly used to answer the "what" questions.

Explanatory research is a study that discovers and reports all facets of a phenomenon under investigation. It studies existing subjects for a good understanding of relationships between variables and answers the "why" questions (Babbie, 2011). It is built on descriptive and explorative type of research. It also gives the reasons why something happens (Neumann, 2013).

Exploratory research is chosen because it will help us to create a quantitative framework and test the feasibility of abstractive text summarisation at the paragraph level.

3.3.1 Research philosophy

A research philosophy is presented as a claim to knowledge, with implications that since scientific research processes point towards the discovery of truths about knowledge (Flick, 2011). Uddin and Hamiduzzaman (2009) indicated that it analyses the social world's entities by viewing it from different perspectives such as reality, scientific truth, nature of knowledge and logic of abstract phenomena. It helps in guiding researchers to view and carefully analyses the areas of their research. The research philosophical convictions include ontology and epistemology (Creswell, 2013).

Ontology seeks to know what constitutes the reality, how society should be viewed and understand (Bracken; 2014; Bhattacherjee, 2012).

Epistemology refers to "theory of knowledge", seeking to understand knowledge, how can it be acquired, and its validity on any subject (Wahyuni, 2012; Krauss, 2005). Bhattacherjee (2012) revealed that epistemology searches for information and facts that can be proved without doubt in your field of research, rather than changeable opinions and situations. That is why this philosophy is the most suitable for this research.

Epistemology philosophy consists of three philosophical positions, which are critical realism, interpretivism and positivism.

3.3.1.1 Interpretivism

Neuman (2013) defines interpretivism as a "systematic analysis of social meaningful action through the direct detailed observation of people in natural settings in order to understand and interpret how people create and maintain their social world". Burke (2007) states that, it helps to understand society, how things are happening and what can possibly happen in the future. The research must be directly observed as experienced by the people. Interpretivism approach makes use of observations, interviews and analysis of existing literature to get a meaningful reality (Myers, 2013, p.39-42).

3.3.1.2 Critical realism

Critical realism tries to understand conflicts produced by cultural, political, social and economic factors (Neuman, 2013).

3.3.1.3 Positivism

Positivism seeks to verify scientific truths through experimental observations, measurements and analysis of the observed phenomena (Babbie, 2011). This approach allows the researcher to acquire knowledge, objectively and independent of the social actors. It is suitable for this study since it allows us to evaluate ROUGE scores of articles summarised.

To obtain a meaningful understanding of the use of ABS at paragraph level. This study was conducted from a positivism standpoint, employing a quantitative method to data collected and analysis using deductive theory-building.

3.3.2 Research approach

Research approach helps the investigator to understand, choose a suitable method for data collection and tools that will help to interpret and analyse this data (Mlitwa, 2011). Research approach is classified into three approaches, known as deductive, abductive and inductive approaches (Neuman, 2013; Saunders, Lewis & Thornhill, 2012). The most used are inductive and deductive approaches.

Inductive approach uses qualitative data collected to create a theory from the analysis of data. This approach does not initially use a framework to inform data collection. Research focus is formed after data has been collected (Flick, 2011). Data analysis may fit existing theory or new theories may be created (Bryman & Bell, 2011). Inductive approach is used in qualitative research, where the absence of theory informing the research process may be beneficial.

Deductive approach forms hypothesis from an existing theory and formulates research approach to test it (Silverman, 2013). It is useful where the investigation is concerned, to examine whether the phenomena observed fit well with expectation, based on the past research (Wiles et al., 2011).
This research used deductive approach to test hypothesis using empirical data (Wilson, 2010).

3.3.3 Research methodology

Research Methodology is a combination of techniques and methods used in a specific research context to carry out research within a specific paradigm (Mlitwa, 2011). Neuman (2013) and Williams (2011) mentioned that research methods can be either qualitative, quantitative or mixed depending on the purpose and the investigation of the research.

Qualitative research is a collection of methods and techniques used to gather and analyse non-numerical data such as texts or images (Harwell, 2011). Qualitative are frequently in a natural setting (workplace of participants). It allows researchers to be involved in the natural experiences of participants as well as be part of the environment of study. Therefore, it allows us to understand the context in which participants in the study address the problem (Creswell, 2013).

Quantitative research is a collection of methods and techniques which involve the collection and analysis of numerical forms of data (Creswell, 2013). This method of enquiry is empirical and includes mostly experiments and surveys. The aim of quantitative research is to measure results on variables which have been studied (Yin, 2013). It will be used in this research to test if the hypothesis stated is true or not.

3.3.4 Research strategy

Harwell (2011) defined research strategy as strategic and details planning as well as the execution of a study. It is used as a guideline to conduct the research. There are many research strategies such as action research, survey, experiments, case studies and grounded theory (Saunders et al., 2009; Oates, 2005, p.33). Quantitative methodology uses mainly surveys and experiments. The research uses experiments to test the hypothesis of data collected.

Experiment is a process used to disprove, support or validate a hypothesis (Bhattacherjee, 2012). It follows a strict design and manipulate variables to produce results that are used to validate the objective of the research (Creswell, 2013). An experiment is used to manipulate and control independent variables, taking note of the final effect of the independent variables on the dependent variable and conduct the experiment with the same experimental design. Further, the experiment is done to find out what happens to something with a specific condition. The researcher tries to control factors that may affect the outputs of the experiments to determine or predict what may happen. The results obtained from experiment are quantitative data in nature.

Surveys consist of sampling a proportion of the population (Check & Schutt, 2012, p.160; Bryman & Bell, 2011). They produce quantitative data that is empirically analyzed. Surveys are to study significant variables between different types of data.

From the above-mentioned, experiment is chosen to text hypothesis of data collected.

Creswell (2013), defines a variable as a measurement of characteristics in a theory. There are two types of variables: independent (input) and dependent (output) variables.

- Independent variables are object/entity that are not changed or influenced by other variables. In practice they are variables researchers conducting experiments tweak to test the hypothesis and introduce output. Examples of independent variables: treatment, manipulated or forecaster variables.
- Dependent variables are outcomes of independent variables.

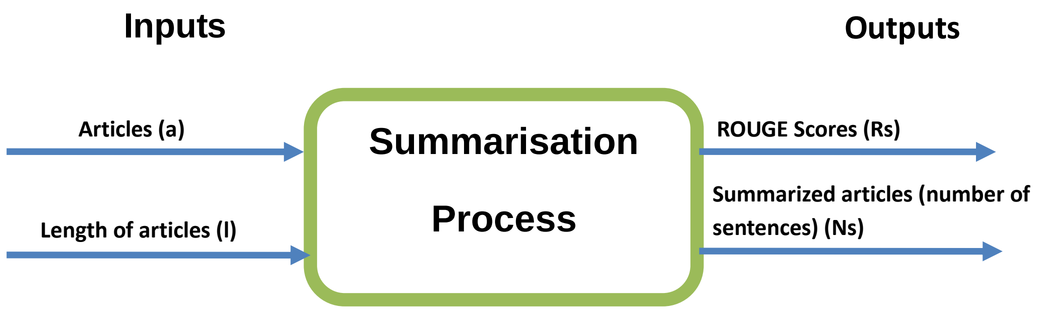


Figure 3.3: Variables definition

According to the Conceptual framework, we defined two main variables as input and output values.

- **Input variables:** articles (a) and length of article (I)
- **Output variables:** Rouge score (Rs) and number of sentences (Ns).

The independent variables are (a) and (I) which will be collected from the open data source like news feeds. However, as output we will get two dependents variables: Rs and Ns.

3.3.5 Data collection

Bhattacherjee (2012) defines data collection as a process of gathering data from participants during a scientific investigation period. It helps the researcher to set boundaries for the study. Quantitative data collection procedures include experiments, surveys and case studies. It helps researchers to get valuable information required to carry out research. Experiment is chosen for this investigation with respect to the research objectives and hypothesis from quantitative data collected (Oates, 2005, p.126-140). In this research data is collected from open news feeds.

3.3.6 Data analysis

Neuman (2013) defines data analysis as the process of converting data (voice, image and text) collected through documents, observations and interviews into meaningful and valuable information for the intended purpose. The purpose of data analysis is to produce an understanding of elements that have collected data. The process involves splitting data into manageable patterns and relationships (Calmeyer, 2011). Data analysis is used to compare the ROUGE scores and the number of sentences of abstractive text summarisation at the paragraph level to sentence level. According to Oates (2005) observation method is used to generate data because it involves senses other than sight (smelling, touching and testing).

3.3.7 Evaluation parameters

ROUGE and BLUE metric are often used to evaluate machine summary compared to human summaries. From the set of ROUGE evaluation metrics, ROUGE-N is the most dominant metric used for text summarisation. Parameters for supervised and unsupervised learning and information retrieval tasks used to measure algorithms performance are ROUGE-N and MEAD. These are common metric used for supervised and unsupervised learning and information retrieval tasks (Hanczar and Nadif, 2018).

The ROUGE metric for evaluation is explained below.

- ROUGE

ROUGE-N is a metric which evaluates the number of n-grams in system summaries that occur in the reference summaries or human summaries.

$$ROUGE - N = \frac{Count\ of\ matched\ n - grams}{Number\ of\ n - grams\ references\ summary}$$

The main assessment metric from ROUGE-N score is Recall, Precision and F-Score.

- Precision

Precision in ROUGE context is the number of overlapping words (sentences) that are in both machine and human summary divided by the number of sentences in machine summary. Precision is calculated according to the formula below.

$$Precison = \frac{Number\ of\ overlapping\ words/sentences}{Total\ number\ of\ system\ summary}$$

- Recall

Recall in ROUGE context is the number overlapping words/sentences that are in both system and reference summary divided by the number of sentences in the reference summary. Recall uses the formula below.

$$Recall = \frac{Number\ of\ overlapping\ words/sentences}{Total\ number\ of\ refence\ summary}$$

- F-score:

F-score is the harmonic average of the recall and precision.

$$F - Score = \frac{2(Precision * Recall)}{Precision + Recall}$$

The range of these metric above are between 0 and 1.

3.4 Ethical considerations

Ethical considerations are fundamentals to guarantee anonymity and confidentiality of participants in carrying out the research study (Babbie, 2011). Ethical clearance has been granted by the company on which this research is based. The research will comply with ethical principles and requirements of the Informatics and Design Faculty of Cape Peninsula University of Technology. In addition, it shall comply with the principles of experimental research. It shall not manipulate the processes of data collection and analysis. The study will use open source software and operating systems as such must comply with the terms and conditions thereof. Documents will come from available public sources. The result of the research will be a contribution to the open source community.

3.5 Chapter summary

Chapter three present the conceptual framework used to describe the concepts and relationships of different variables of abstractive text summarisation using RNN at the paragraph level. Then clarifies with the use of research design how the parts of the research process are connected and organized in a way to lead answers the hypothesis of the research.

CHAPTER FOUR

EXPERIMENT PLANNING, SETUP AND IMPLEMENTATION

4.1 Introduction

This chapter focuses on describing the plan used in executing the methodological protocol in chapter three and discusses in detail how the experiment is made. It describes the experimental goals to find out what the research is trying to answer, experimental materials, tasks and procedures implemented to carry-out the research. It talks about how news articles was stored in the database, the parameters or settings for the algorithm. It also used a diagram to explain the experimental setup and responses on the research hypothesis stated in the first chapter.

4.2 Experimental goals

This section describes the complete manipulation of the experimentation.

- **Goal 1**: The main goal is to design and develop an abstractive text summarisation using RNN at the paragraph level application software. This will combine a machine learning algorithm with an attention encoder-decoder to encrypt and decrypt files during the summarisation process, rather than human summarisation documents, reduce time consuming and summarisation costs.
- **Goal 2**: Training ML algorithm to summarised articles such online DUC-2004.
- **Goal 3**: Compare the results of abstractive text summarisation using recurrent neural networks at the paragraph level with sentence level.

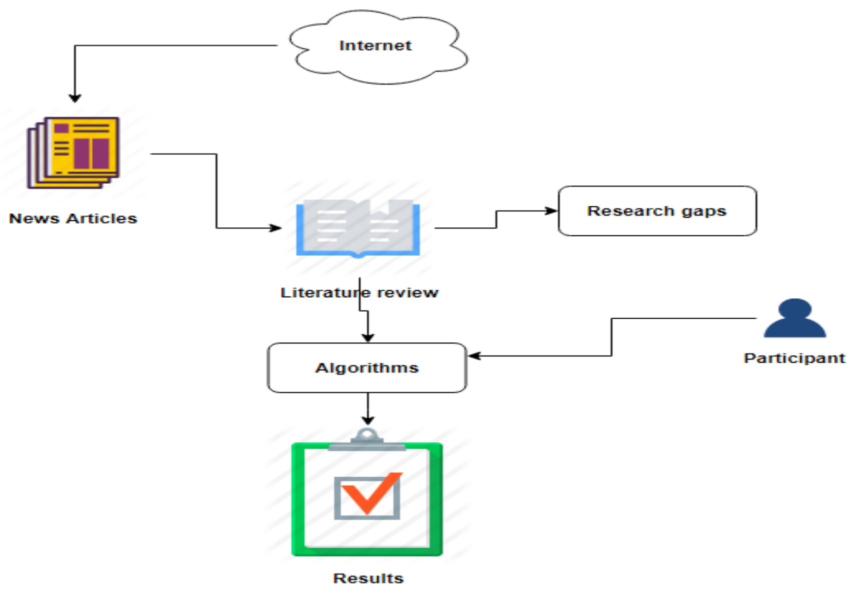


Figure 4.1: Overview of the research goal

Articles are gathered from the internet and fed into algorithms that produce summarised documents. The participant manipulates the variables of the algorithm and identifies possible accuracy of the results. The literature identifies possible research gaps. As from tests done I use Figure 4.1 to generate the results of the experiments.

4.3 Participants

In this study, the participant is the researcher involved in the implementation of the algorithm as well as the provider of data collected from DUC-2004 online dataset that allows to do analysis. The University provides all the resources to carry-out the research and the simulation of the experiment including the availability of Cape Peninsula University of Technology (CPUT) laboratory to access resources.

Ethical methodology described in Chapter three in section 3.4 ensures that we will not violate any principle of the research participants. Further, it helps to protect the Institution and the research against potential legal implications from any behavior that may be deemed unethical during the experiment process.

To ensure ethical treatment of the research subject and validity, the data was collected from the experiment.

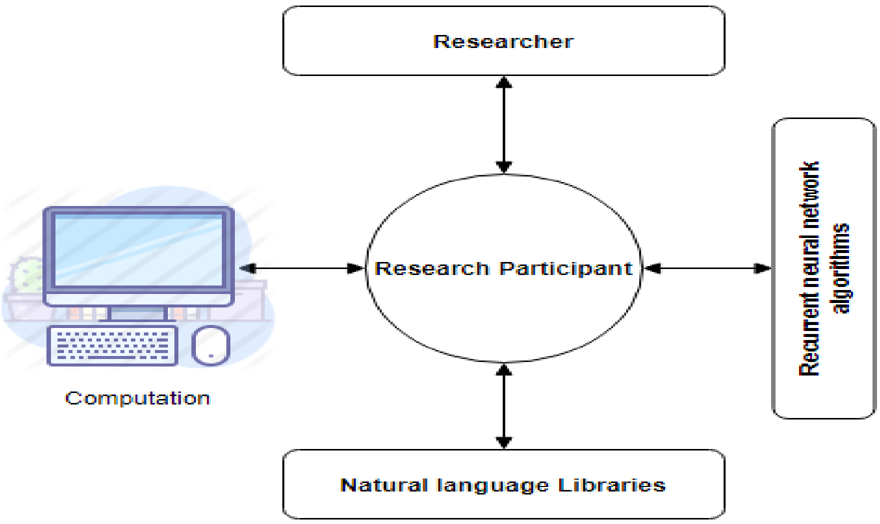


Figure 4.2: Research place and participants

The experiment had no participants beside the researcher (CPUT student). In that regard there was no need for ethical clearance or consideration before undertaking the research. Figure 4.2 present algorithms and libraries used in this research to run the experiment.

The researcher worked at CPUT Laboratory, used the resources afforded by the University within the laboratory to perform the research experiment by developing an application that can summarise articles.

4.4 Experimental materials

Experimental materials and equipment used in this study are described. The characteristics that can impact the results are described as detailed and precisely as possible below to allow the reader to understand materials used during the experiment.

- HP Intel i5 8^{th} Generation with built-in 4GB RAM, extended 16GB RAM, 500GB HDD, Ubuntu 18.04.1 or Windows 8 Operating System.
- External 2TB HDD used for abstractive text summarisation experiment and storage of dataset requirements.
- Operating system is the open source-based Linux or any compatible operating system.
- Python programming language was used for development of application.
- IntelliJ IDEA 2019.3.4 tools to implement python language.
- ML approach was implemented (Unsupervised Learning) for text summarisation.
- An algorithm that takes single stories from the database and generates a summary of the story.
- A suitable evaluation model pyrouge (version py-rouge-1.1) package which provides the measure Recall, Precision and F-score.

4.5 Tasks

Tasks performed during the experiment were described in detail to allow possible replication of the experiment without consultation of the authors.

Guided by the conceptual framework in Chapter three at section 3.2, the research tasks were structured through the following steps to complete the experiment:

- The stories are collected from websites and manually stored in a cockroach DB database.
- Load articles one at a time into the abstractive summarisation application.
- Generate summarised text files.

- Obtain the ROUGE scores and number of sentences of texts summarised.
- Compare the ROUGE scores and number of sentences of text files summarised at paragraph level to sentence level.

4.6 Hypothesis, parameters and variables

To conduct an experiment, it is important to determine and understand the different variable groups (independents and dependents variables) and hypothetical relationship between them. Independent variables in this investigation are each article and its length; dependent variables are ROUGE scores of an article and its number of sentences determined in chapter three section 3.3.4. The hypothesis of the research is discussed in chapter one section 1.4.

4.7 Experimental design

It describes in detail how the process is performed during the experimental stage. It explains the conceptual framework in chapter three Section 3.2:

- **Stage 1** is the summarisation of text files from the online dataset selected.
- **Stage 2** is the comparison of algorithms from the results obtained.

4.8 Procedure

The procedure describes the setting of the experiment. Then details of the data collection and method is described above. The articles were collected from online DUC 2004 dataset. The stories with adverts and pictures were cleaned so that only text documents remain. Then data was stored in a database. Further, were passed through the process of summarisation to measurement of ROUGE scores of summarised texts one at a time.

4.9 Deviation from the plan

Each experiment has a risk of deviation from the setup plan. An alternative plan should be put in place to accommodate such eventualities. Data can be collected from different online datasets such as Gigaword, CNN/Daily Mail**,** MSR, NLPCC2017, paste documents in a text file and store in another database such as Cassandra or PostgreSQL. Programming language other than Python can be used to realize this experiment. Another metric of measurement such as MEAD can also be used.

4.10 Experimental diagram

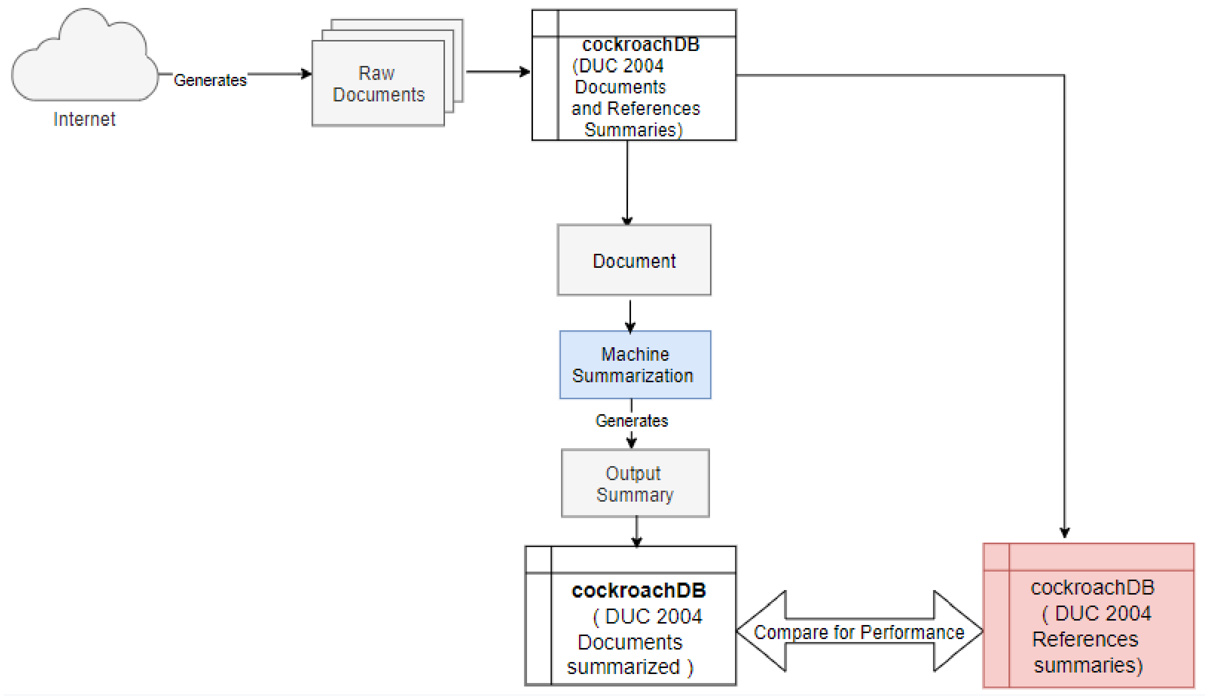


Figure 4.3: Experimental diagram

The experimental diagram in figure 4.3 above was run on the windows machine using Ubuntu 18.04.1. The processor used is an i7-8550U CPU @ 1.80GHz, 4 Cores with 8GB of RAM.

The input documents of the experiment came from the internet DUC database. Our focus is on DUC 2004 raw documents. From the documents collected, each document (article) is used to generate the machine summary and compare the performance with the alternative set of four reference summaries provided by the DUC 2004 online database.

4.11 Parsing documents to the algorithm

The collection of the articles was an unsupervised process because the algorithm used to summarise the articles was totally unsupervised and the data collected stored in cockroach DB database. The complete steps of summarisation are described in figure 4.8.

The articles collected from the internet were singled documents in plain text. Documents cannot be parsed to the RNN algorithm in plain text. It needs to convert into a numerical value or document vector to process the RNN algorithm. There are many techniques used for numerical representation of documents to vector but much of them produce low performance. The numerical representation techniques of documents used for the experiment are discussed below.

4.12 Document vector representation techniques

Document vector representation aims to represent unstructured text document by a numerical vector such that the similarity between vectors and documents can be mathematically computed by different kernels. There are many documents vector representation techniques, only Word2Vec and Doc2Vec are discussing in this investigation to address the disadvantages of bag-of-words document representation.

4.12.1 Word2Vec

Word2vec is a model used to transform and represent words into vector space. It captures the semantic relationship, and then models their word's context such as synonyms, analogies and antonyms etc. Each word can be assigned a vector in the space and a word vector can be several hundred dimensions. Words are converted to vector representation to perform numerical value operation rather than on text (Landthaler et al., 2017; Mikolov et al., 2013a). The resulting word vector dimension enables to save storage and computational resources. Word2Vec is a versatile static model, which cannot be dynamically optimized for specific tasks. Word2vec makes use of two algorithms known as Continuous Bag-Of-Words (CBOW) and Skip-Gram (SG) to produce representation of words.

4.12.1.1 Continuous Bag-Of-Words

The main idea of CBOW is to guess target words from the surrounding word context. This technique produces poor performance, due to the non-consideration of word order (Meyer, 2016; Mikolov et al., 2013a). Figure 4.4 below is a sample of CBOW.

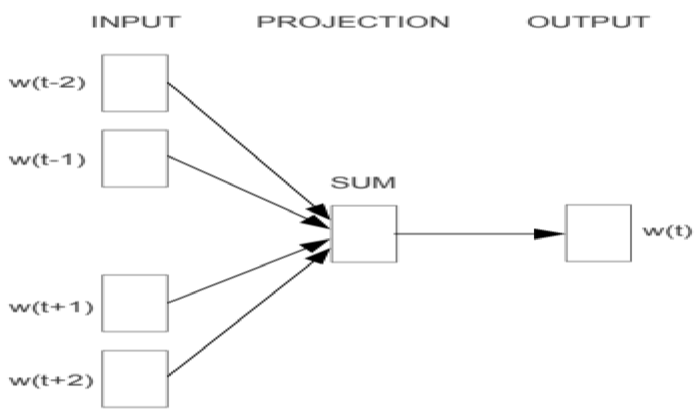


Figure 4.4: Example of CBOW (Mikolov et al., 2013a)

4.12.1.2 Skip-Gram

Skip-Gram (SG) algorithm is used to predict surrounding context words from the target words. SG is a NN model trained on a large vocabulary corpus to perform tasks (Gupta et al., 2019; Mikolov et al., 2013a). The weight of the hidden layer is the vector of the words. When the neural network is given an input word from a sentence, it looks at the nearby words, picks one randomly and computes the probability of the word being in the vocabulary. Figure 4.5 shows an example of SG.

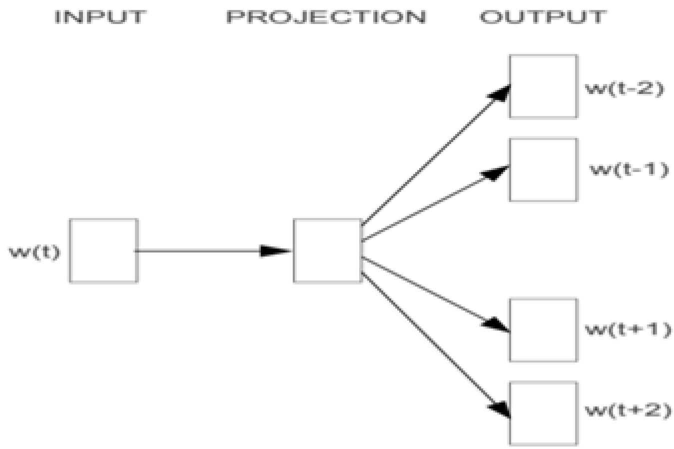


Figure 4.5: Example Skip-gram model (Mikolov et al., 2013a)

CBOW and Skip-gram are standard models for obtaining word embedding based on word-context pair information. According to (Mikolov et al., 2013b), CBOW does better on syntactic (semantic) relationship task, while Skip-gram approach is slower but produces best overall performance in infrequent words. On sentence completion, training a Recurrent Neural Network Language Model (RNNLM) starting with skip-gram word vectors performs best.

4.12.2 Doc2vec

Doc2vec was inspired from Word2vec. Document vector also known as paragraph vector is a method that learns a vector representation of an input text (sentences, paragraph(s) or whole document) regardless of a variable length. This method aims to predict the next word in a paragraph given the concatenation of the actual paragraph vector and various word vectors. The first step is learning the word vectors to use them later for inferring paragraph vectors.

Le and Mikolov, (2014) propose the word vectors for learning distributed words that was inspired by works related to models known as neural language models representation (Mikolov et al., 2013a). Doc2vec is a simple technique, easy to use and known for producing good results in capturing the semantics of paragraphs. It also performs robustly even when trained using large corpora. It is embedded Paragraph Vector Distributed Memory (PV-DM) and Paragraph Vector Distributed Bag-Of-Words (PV-DBOW).

4.12.2.1 Paragraph vector distributed memory

PV-DM is a method for generating vectors from an input text or document. PV-DM was developed from CBOW model. It uses a target word to predict a context and a sliding window to create a vector of a whole paragraph. A Softmax function predicts context for all the words in the sentence as a window slide creates word embedding. This embedding is averaged / concatenated. From each paragraph one vector is created and for each word one vector is created. These vectors created are trained using stochastic gradient descent. The gradient error is calculated to update the parameters in the model using a random fixed length context (Le and Mikolov, 2014).

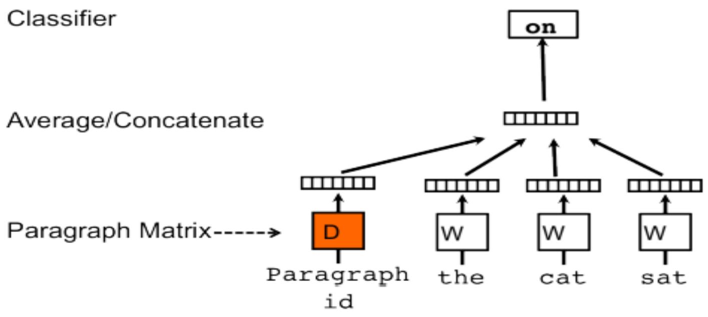


Figure 4 6: PV-DM model (Le & Mikolov, 2014)

4.12.2.2 Paragraph vector distributed bag-of-words

PV-DBOW uses a word to predict a context. It is faster than Word2vec and consumes less memory. It also updates the parameter and works the same way as Skip-gram. Doc2vec produces two vectors, one by PV-DM and another by PV-DBOW (Le and Mikolov, 2014). PV-DM achieves better results compared to PV-DBOW. PV-DBOW model of Doc2Vec is used in this experiment.

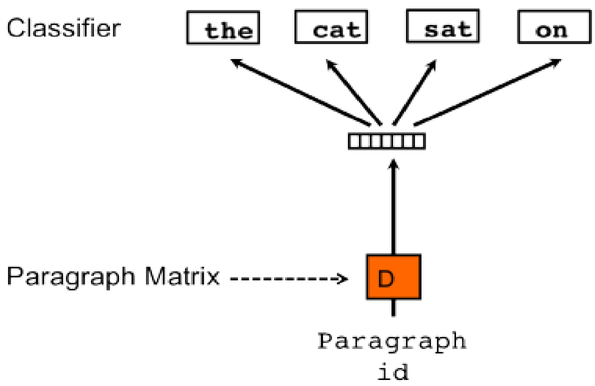


Figure 4.7: PV-DBOW model (Le & Mikolov, 2014)

Doc2Vec performs well with small/large corpora compared to Word2Vec. Furthermore, you train in Word2vec to find word vectors and run similar vectors between words, while in Doc2Vec if you tag a document you also tag vectors and vice-versa.

4.13 Summarisation model

The RNN model and attentional-based mechanism is used in this work. This model fed the tokens of the article into the encoder (RNN layer). The decoder (RNN layer) gets the embedding (Doc2vec) of the preceding word (paragraph) on each time interval. Attentional is computed as a weight of the sum of a set of encoder hidden states, dependent of the current decoder hidden state (Bahdanau et al., 2015). The attention distribution tells the decoder where to look to produce the next word. A pointer generator is used to manage out of vocabulary words and accurate mistakes. On each time interval, a probability is generated from the context vector, the decoder input and state.

4.14 Performance

The algorithm is fed with the same set of data from DUC2004 and run as much as possible to reach the maximum level of satisfaction. The output of the algorithm will be the text summarised. The performance of the algorithm will be measured in terms of runtime, length of text and ROUGE scores. The metric used to access the performance of ROUGE scores are Recall, Precision and F-score.

4.15 Data

The standard paragraph summarisation evaluation is associated with DUC 2004 shared tasks (Over et al., 2007). In this experiment, data collected consists of fifty articles from the Press Wire services and New York Times, each pair with four human reference summaries. To make only recall evaluation unbiased to length, the output length is cut-off to 100 characters and for shorter summaries no bonus is given. There are various versions of ROUGE that match several lengths. The DUC evaluation in our experiment use ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-L and ROUGE-W.

4.16 Parameters

The parameters are settings that are adjusted on the algorithm. The RNN model has 256 dimensional states with pre-trained and trained word embedding from data. We have used a vocabulary size of 32128 tokens. Then use T5Tokenizer to generate pre-trained word embedding. The validation set loss was used for premature stopping. The summaries are created using beam search algorithm with size five while texting. The standard attention was reused as copy. The loss of the sequence is divided by the sum of tokens in it. The first hidden state of the decoder was calculated by an additional layer using the last hidden state of the encoder as input.

4.17 Summarisation steps at paragraph level

Figure 4.8 below illustrates summarisation steps of the algorithm at the paragraph level.

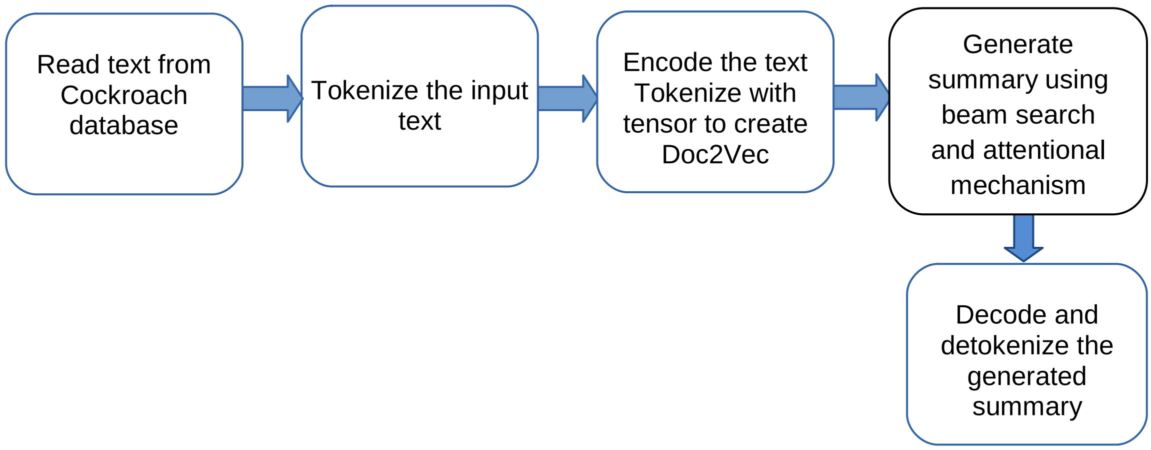


Figure 4 8: Summarisation steps

4.18 Conclusion

Chapter four expanded on the methodology described in chapter three. It discussed the experimental goals, the participants' roles, experimental materials, tasks, procedures that were implemented to carry-out the research. It discussed how news document texts from DUC 2004 were converted into a suitable form for the algorithm to consume. It uses the diagram in figure 4.3 to explain how the experiment was run step by step and outlines the research process. It gives the parameters or settings for the algorithm. It also explained how the deviation from the plan can be controlled.

CHAPTER FIVE

FINDINGS AND DISCUSSIONS

5.1 Introduction

This chapter discusses the outcomes and findings of the experiments. The result is the output summary of a DUC 2004 article, then its ROUGE scores. The results are given in the form of tables and graphs.

5.2 Paragraph and sentences per document

Table 5.1: Paragraphs and sentences of documents

DUC 2004 data set parameter	Words per document	Sentences per document	Paragraphs per document
Maximum	14299	655	605
Minimum	3102	153	152
Average	5870	291	166

Table 5.1 shows the number of minimum and maximum words, sentences and paragraphs from the fifty DUC 2004 documents used during the experiment.

5.3 Runtime in seconds

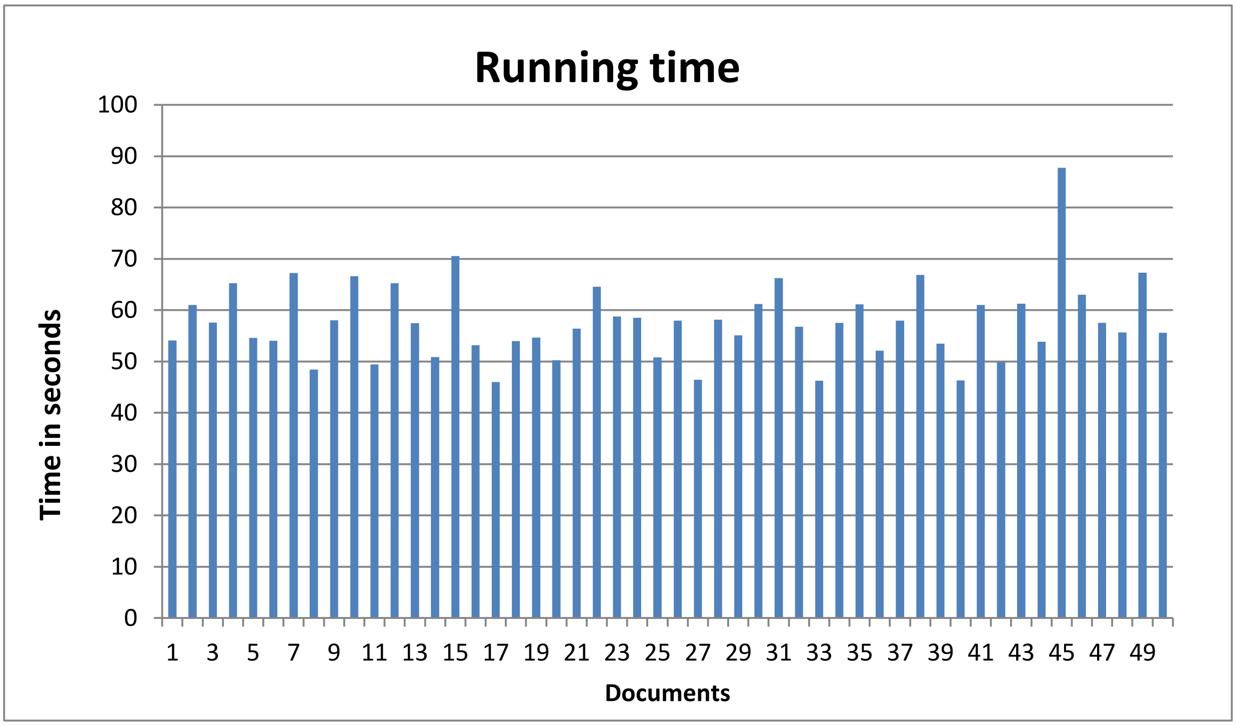


Figure 5.1: Execution time of documents in seconds

The graph in figure 5.1 above indicates the time (in seconds) that it takes for the algorithm of each document to produce a summary. The best time it takes for the execution of the algorithm was 46.27 seconds by the document number 33 and the worse time was 87.71 seconds by the document 45.

5.4 Precision, Recall and F-score Performance

The performance of the algorithm is evaluated using ROUGE-N based on coherence and text readability of the summary. The higher ROUGE score metric indicates a higher match of summaries produced from original texts.

5.4.1 ROUGE-1 results for all metrics

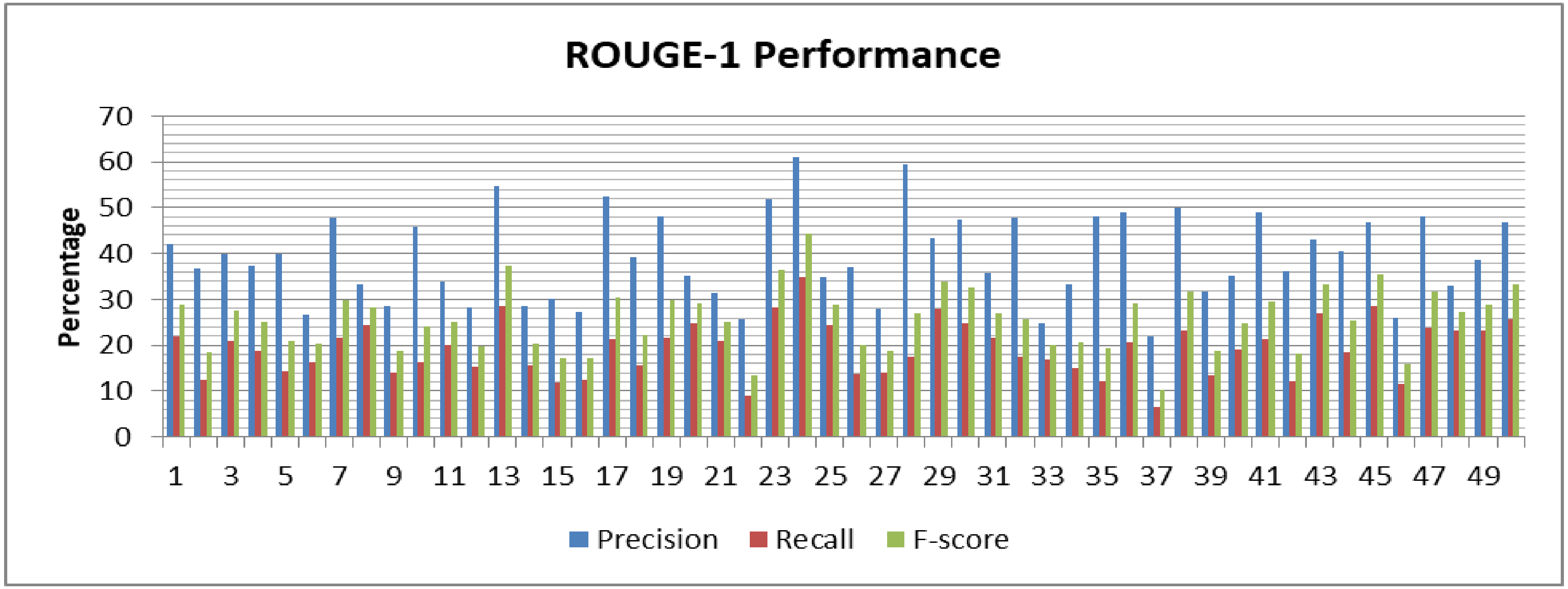


Figure 5.2: ROUGE-1 metrics performance

The graph in figure 5.2 presents the ROUGE-1 metrics performance.

The best Precision is achieved by document 24 and the worst Precision is by document 37. The best Recall is achieved by document 24 and the worst Recall is by document 37. The best F-Score is achieved by document 24 and the worst F-score is by document 37.

5.4.2 ROUGE- 2 results for all metrics

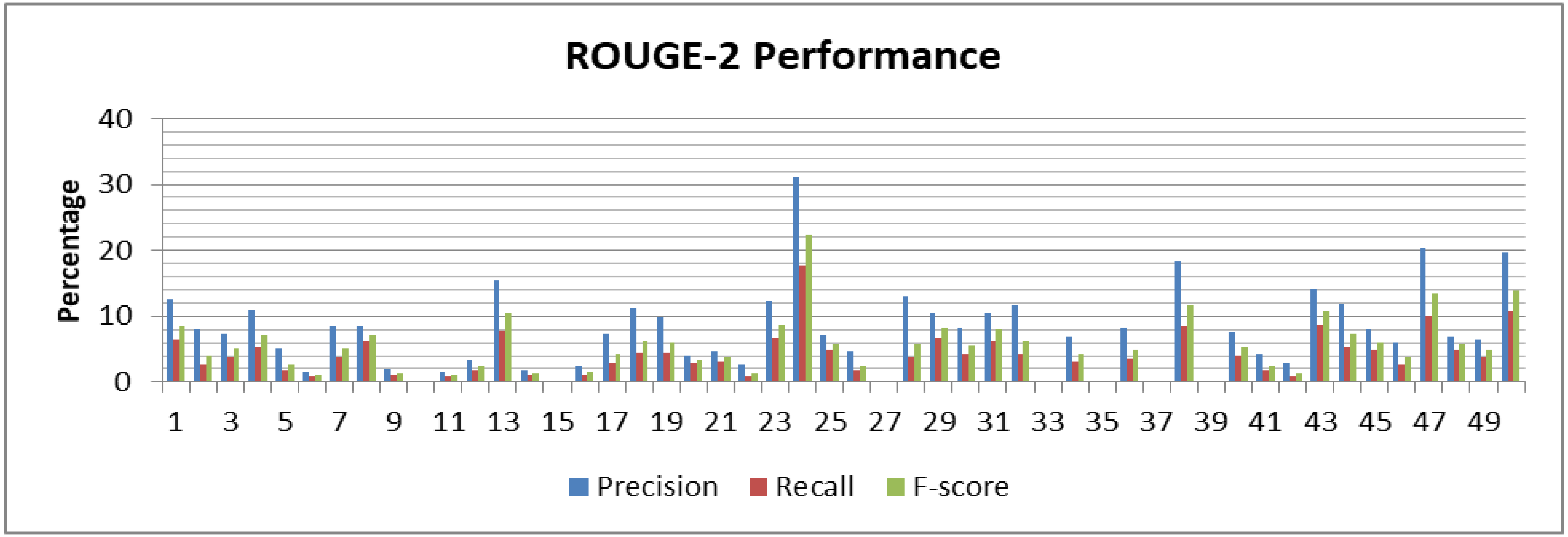


Figure 5.3: ROUGE-2 metrics performance

Figure 5.3 above presents the ROUGE-2 metrics performance.

The best Precision is achieved by document 24 and the worst Precision is by documents 10,15,27,33,35,37,39. The best Recall is achieved by document 24 and the worst Recall is by documents 10,15,27,33,35,37,39. The best F-Score is achieved by document 24 and the worst F-score is by documents 10,15,27,33,35,37,39.

5.4.3 ROUGE -3 results for all metrics

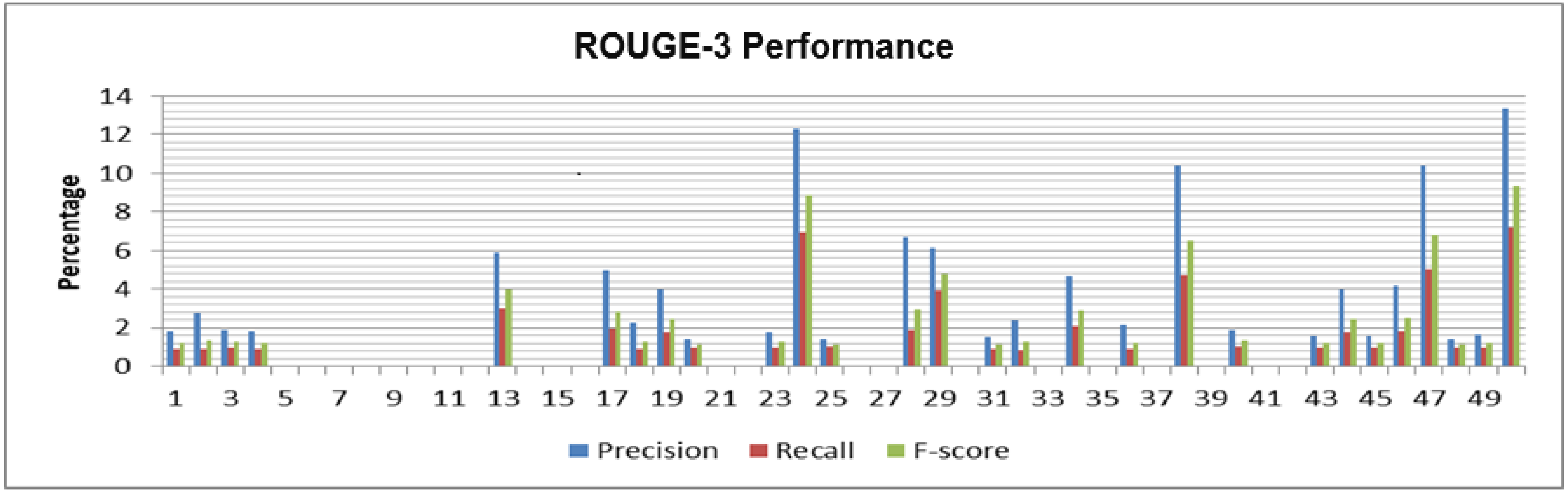


Figure 5.4: ROUGE-3 metrics performance

Figure 5.4 above presents ROUGE-3 metrics performance.

The best Precision is achieved by document 50 and the worst Precision is by documents 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 21, 22, 26, 27, 30, 33, 35, 37, 39, 41, 43. The best Recall is achieved by document 50 and the worst Recall is by documents 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 21, 22, 26, 27, 30, 33, 35, 37, 39, 41, 43. The best F-Score is achieved by document 50 and the worst F-score is by documents 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 21, 22, 26, 27, 30, 33, 35, 37, 39, 41, 43.

5.4.4 ROUGE–L results for all metrics

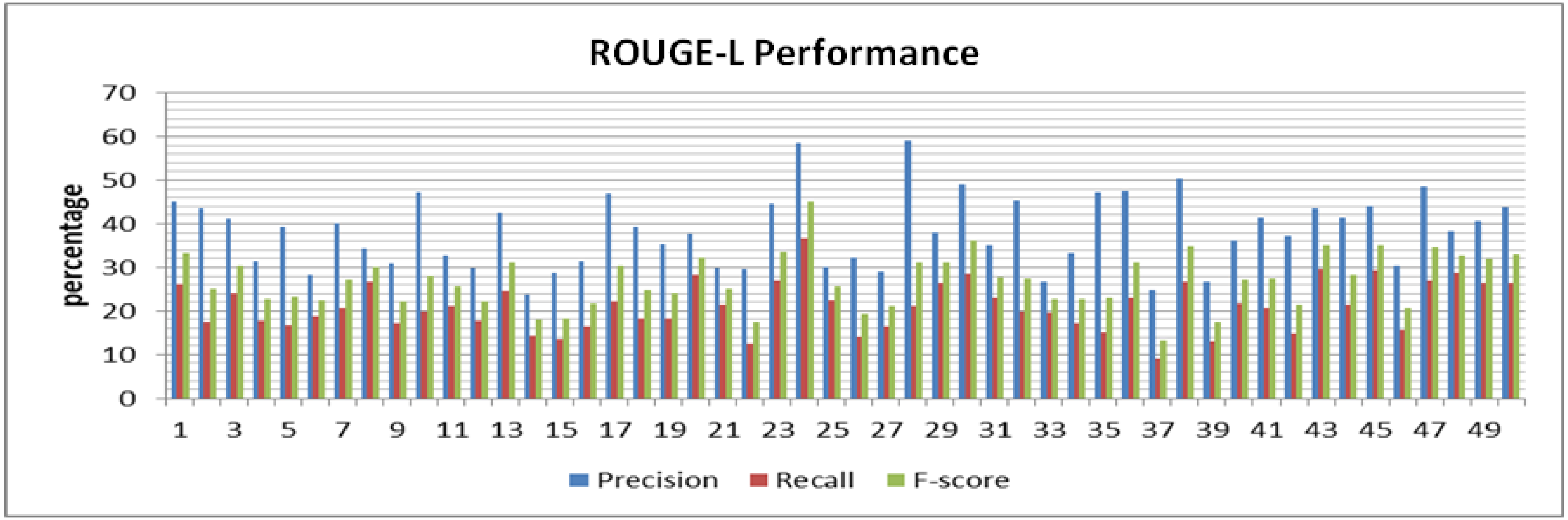


Figure 5.5: ROUGE-L metrics performance

The graph in figure 5.5 presents ROUGE-L metrics performance. The best Precision is achieved by document 28 and the worst Precision is by document 14. The best Recall is achieved by document 24 and the worst Recall is by document 37.

The best F-Score is achieved by document 24 and the worst F-score is by document 37.

5.4.5 ROUGE-W results for all metrics

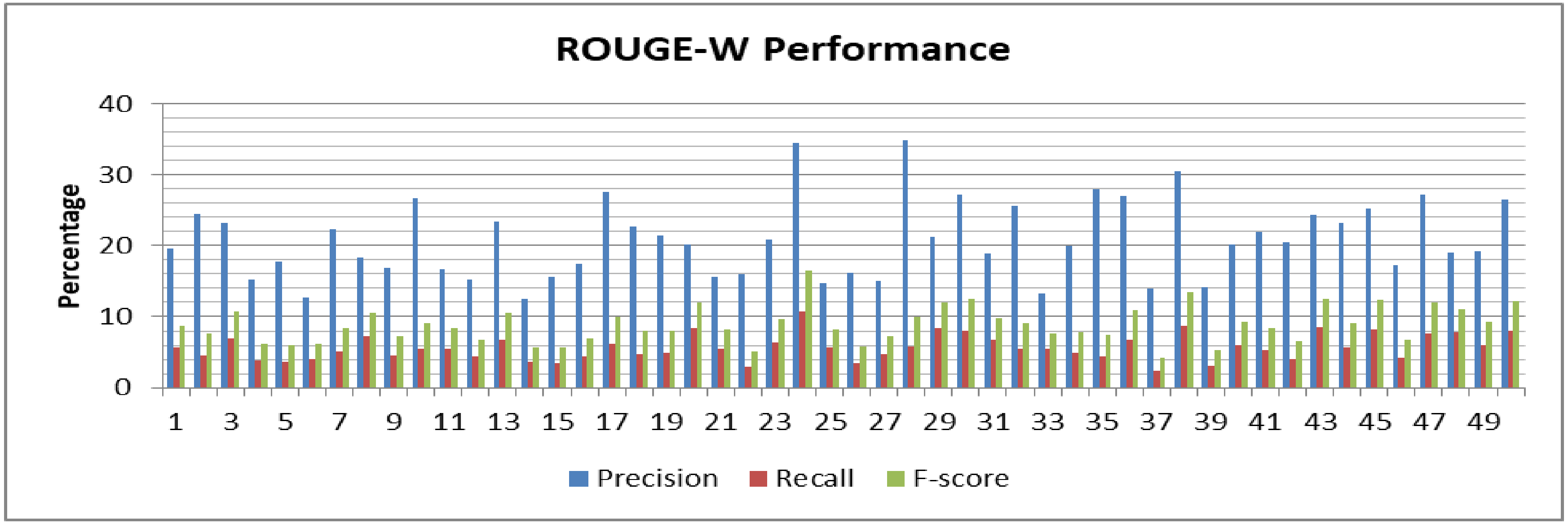


Figure 5.6: ROUGE-W metrics performance

The graph in figure 5.6 presents ROUGE-W metrics performance.

The best Precision is achieved by document 28 and the worst Precision is by document 14. The best Recall is achieved by document 24 and the worst Recall is by document 22. The best F-Score is achieved by document 24 and the worst F-score is by document 37. From the results of the all the ROUGE-N scores graph obtained above, we can deduce the performance of the algorithm of the ROUGE-score details in section 5.5.

5.5 ROUGE score evaluation

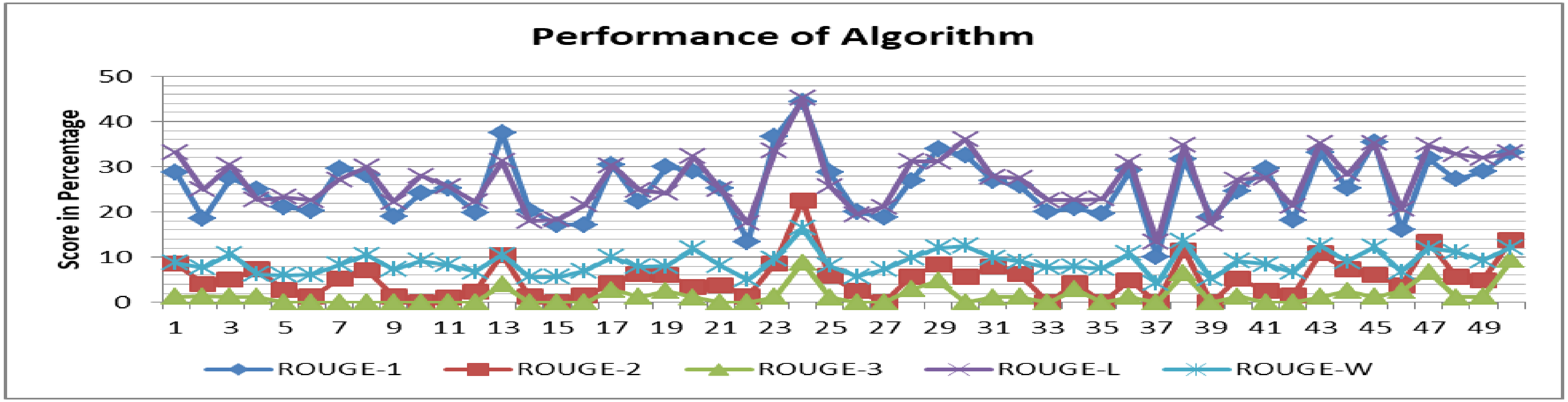


Figure 5.7: Algorithm performance

Analyzing the ROUGE scores, it is clearly shown in figure 6.7 that ROUGE-1 and ROUGE-L achieved the better scores than ROUGE-2, and ROUGE-2 achieved better scores than ROUGE-W. ROUGE-3 achieved the worst score of the algorithm. Table 6.2 below gives the summary of ROUGE-N of the algorithm.

Table 5.2: Experiment result on main summary task on ROUGE metric

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-L	ROUGE-W
Best	44,44	22,5	9.36	45.15	16.46
Mean	25,55	5,06	1.52	26.9.36	8.85
Worse	10,14	0	0	13.35	4.2

5.6 Results and findings

The results answer the hypothesis stated in chapter one such as:

- Abstractive text summarisation using RNN at paragraph level produces different number of sentences and improves ROUGE scores compared to sentence level abstractive text summarisation or not?

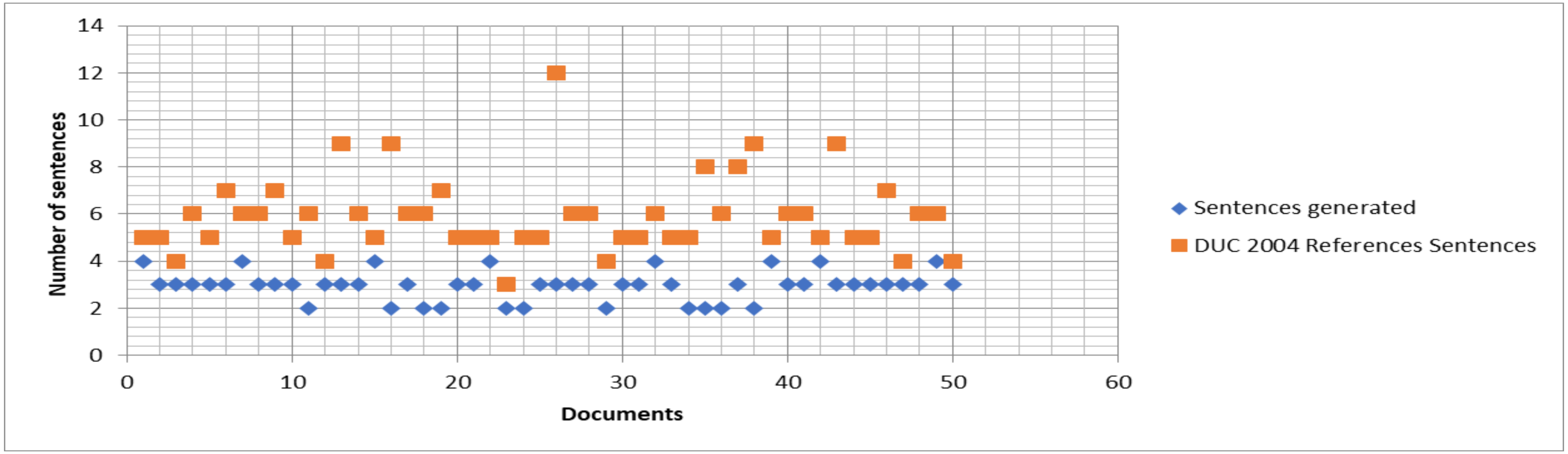
According to Rush et al. (2015) the abstractive text summarisation at sentence level best ROUGE scores on DUC 2004 and Gigaword as follows in table 6.3. Further, the experiment results of abstractive text summarisation at the paragraph level ROUGE scores are also given in table 6.3

Table 5.3: Comparison of experimental result on the summary task ROUGE metrics of abstractive text summarisation using RNN at sentence level and the paragraph level

		DUC 2004			Gigaword		
	Model	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
RNN at sentence level	**ABS**	26,55	7,06	22,05	30,88	12,22	27,77
	ABS+	28,18	8,38	23,81	31,00	12,65	28,34
RNN at Paragraph level (Experiment result)	**ABS**	44,44	22,50	45,15			

The results of the experiment prove that abstractive text summarisation using RNN at paragraph level perform better than abstractive text summarisation using RNN at sentence level. It produces better ROUGE scores than DUC 2004 and Gigaword at sentence level. Furthermore, it surpasses and produces the state-of-art at DUC 2004 of all papers selected from abstractive text summarisation in table 2.4**.**

Table 5.4: Compare number of sentences produced between the algorithm and the alternative reference summaries provided by DUC 2004



The number of sentences produced from this experiment is mainly two, three or four sentences from the documents summarised with the maximum of seven four words from the documents summarised. This experiment produces at least one sentence less than the alternative 4 reference summaries provided by DUC 2004 dataset. Rush et al. (2015) proved in their examples given with Gigaword that the generated summaries of abstractive text summary using RNN at sentence level have the same number of sentences with the reference summaries using RNN at sentence level.

5.7 Further direction and recommendation

The experiment runs successfully. However, it would have been desirable to have more data and have more computation power to test the algorithm. No possible settings run could be achieved. It would have been interesting to run this data on other datasets such as: Gigaword and Daily Mail. Only 50 articles were available on DUC-2004 online dataset and their reference summaries, no server with more dedicated resources available to run efficiently, no time to search online Gigaword and Daily Mail dataset.

The abstractive text summarisation using RNN at paragraph level improved the grammaticality of text summarised and the ROUGE score compared abstractive text summarisation using RNN at sentence level. Further research still needs to be done to improve the ROUGE score since the generated summary is still far below human summaries.

5.8 Conclusion

This chapter discusses the results and finding from this experimental study. DUC 2004 online dataset articles were used to run successfully the experiment. Furthermore, ROUGE-1, ROUGE-2, ROUGE-L were utilized to evaluate the quality of summaries. The experiment showed that the highest values of ROUGE-1, ROUGE-2 and ROUGE-L were obtained in abstractive text summarisation with RNN at paragraph level, with values of 44.44, 22.50, and 45.15 respectively. The results shown in the table 5.3 and table 5.4 presented the comparison between abstractive text summarisation using RNN at sentence level and at paragraph level. The challenges faced during summarisation process were unavailability of more datasets and lack of server with dedicated resources.

CHAPTER SIX

CONCLUSION AND FUTUR DIRECTION

The research was motivated by the need to manage the vast number of articles on the internet. Users are presented with a vast number of documents mostly similar. The huge number of information and duplication makes it difficult for users to get information they want with the least amount of time. The reading rate of users, mainly researchers has not changed, yet information continues to grow daily. Duplication of information is annoying to users because they must sift through hundreds of pages, to update themselves on a subject. There is a need to summarise documents.

Literature has shown that abstractive text summarisation is less investigated and widely used today for the task of summarisation. It generates dynamic paraphrases and produces natural summaries compared to extractive summarisation. Research has shown the competitive performance of RNN using the attention mechanism in machine translation. A literature review conducted to align the research with other past studies shows that abstractive text summarisation, using RNN with attention mechanism at sentence level produces accurate abstractive summaries. The research on abstractive text summarisation using RNN at paragraph level on the same set of parameters and constraints at sentence level, still needs to be done.

An experiment was conducted to answer the research hypothesis and goals. The experiment compared abstractive text summarisation using RNN at paragraph level and abstractive text summarisation using RNN at sentence level. Fifty new articles were collected from DUC 2004 online database. The settings have been discussed in chapter four. The ROUGE scores performance was accessed using Precision, Recall and F-score metrics. The results obtained from the experiment show that abstractive text summarisation using recurrent neural networks at paragraph level, produce better quality summaries, ROUGE scores and less sentences than abstractive text summarisation using RNN at sentence level. The experiment also shows no significant difference in performance that can be obtained by increasing the number of text documents in the experiment. As next step, we would like enhance the grammaticality of generated summaries at paragraph level with datasets such as Gigaword and CNN/Daily Mail, as well to investigate abstractive multi-documents summarisation using RNN at paragraph level.

REFERENCES

Abbas, H., Elsied, M., & Salim, N., 2013. Automatic abstractive summarization a systematic literature review, 54(3).

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. T*ext summarization techniques: a brief survey*.

Allen, J., 1995. Natural language understanding. Pearson.

Alpaydin, E., 2014. *Introduction to Machine Learning*. MIT Press.

Aveyard, H., 2014. *Doing a literature review in health and social care: A practical guide*. McGraw-Hill Education (UK).

Ayodele, T. O. 2010. Types of machine learning algorithms. In New advances in machine learning. InTech.

Babbie, E., 2011. *The basics of social research*. 6th ed. M. Kerr, S. Dobrin, R. Jucha, N. Bator, & J. Chell, eds. Canada: Wadsworth, Cengage Learning.

Bahdanau, D., Cho, K. and Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. ICLR2015

Barzilay, R. and Elhadad, M., 1999. Using lexical chains for text summarization. Advances in automatic text summarization, pp.111-121.

Bhattacherjee, A., 2012. Social science research: Principles, methods, and practices, 2nd edition.

Bonissone, P.P., 2015. Machine learning applications. In Springer Handbook of Computational Intelligence (pp. 783-821). Springer, Berlin, Heidelberg.

Bracken, S. 2014. Discussing the importance of ontology and epistemology awareness in practitioner research. *Worcester Journal of Learning and Teaching*, (4): 1–8.

Bryman, A., and Bell, E. 2011. Business Research Methods (3rd ed.) Oxford: Oxford University Press.

Brynjolfsson, E. and Mitchell, T., 2017. What can machine learning do? Workforce implications. Science, 358(6370), pp.1530-1534.

Bullinaria, J.A., 2013. Recurrent neural networks. *Neural Computation: Lecture*, *12*.

Burke, M.E. 2007. Making choices: research paradigms and information management: Practical applications of philosophy in IM research. Library review, 56(6): 476–484.

Calmeyer, T., De Kok, C., Hardy, J., Rogers, J.A. and Taljaard, L., 2011. How to succeed in your master's and doctoral studies: A South African guide and resource book, Johann Mouton: Book review. New Voices in Psychology, 7(2), pp.148-152.

Cavana, R.Y., Delahaye, B.L. and Sekaran, U., 2001. Applied business research: Qualitative and quantitative methods. John Wiley & Sons Australia.

Celebi, M.E. and Aydin, K. eds., 2016. Unsupervised Learning Algorithms. Springer International Publishing.

Check, J., and Schutt, 2012. Survey research. In J, Check & R.K.Schutt (Eds.). Research methods in education. (pp.159-185). Thousand Oaks, CA: Sage Publications.

Cho, K., Courville, A. and Bengio, Y., 2015. Describing multimedia content using attention-based encoder-decoder networks. IEEE Transactions on Multimedia, 17(11), pp.1875-1886.

Chopra, S., Auli, M. and Rush, A.M., 2016. Abstractive sentence summarization with attentive recurrent neural networks. *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 93-98).

Condori, R.E.L. and Pardo, T.A.S., 2017. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. Expert Systems with Applications, 78, pp.124-134.

Creswell, J.W. 2013. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Dalal, V.and Malik, L.G., 2013. A survey of extractive and abstractive text summarization techniques. In Emerging Trends in Engineering and Technology (ICETET), *2013 6th International Conference* on (pp. 109-110). IEEE.

Das, D. and Martins, A.F., 2007. *A survey on automatic text summarization. Literature Survey for the Language and Statistics II course at CMU*, 4, pp.192-195.

Donalek, C., 2011, April. Supervised and unsupervised learning. In Astronomy Colloquia. USA.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate.

Edmundson, H.P., 1969. New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2), pp.264-285.

Elena Lloret and Manuel Palomar, 2012. *Text summarization in progress: A literature review. Artificial Intelligence review* 37(1), 1-41, 2012-Springer.

Elsied, H.A.M. and Salim, N., 2013. Automatic abstractive summarization a systematic literature review. *Journal of Theoretical & Applied Information Technology*, 54(3).

Erman, N., Korosec, A. and Suklan, J., 2015. Performance of selected agglomerative hierarchical clustering methods. Innovative Issues and Approaches in Social Sciences, 8, pp.180-204

Exner-Stöhr, M., Kopp, A., Kühne-Hellmessen, L., Oldach, L., Roth, D. and Zimmermann, A., 2017. The potential of Artificial Intelligence in academic research at a Digital University. Digital Enterprise Computing (DEC 2017).

Flick, U., 2011. *A beginner's guide to doing a research project*. Introducing Research.

Gaikwad, D.K. and Mahender, C.N., 2016. A review paper on text summarization. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(3), pp.154-160.

Gambhir, M. and Gupta, V., 2017. *Recent automatic text summarization techniques: A survey. Artificial Intelligence Review*, 47(1), pp.1-66.

Gardner, M.W. and Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric environment, 32(14-15), pp.2627-2636

Gary, E. & William, A.R. 2017. How to choose machine learning algorithms | Microsoft Docs. Microsoft.

Gatt, A. and Krahmer, E., 2018. Survey of the state of the art in natural language generation: core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61, pp.65-170.

Genest, P.E. and Lapalme, G., 2012. Fully abstractive approach to guided summarization. *In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2* (pp. 354-358).

Ghahramani, Z., 2001. An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(01), pp.9-42.

Gholamrezazadeh, S., Salehi, M.A. and Gholamzadeh, B., 2009, December. A comprehensive survey on text summarization systems. *In Computer Science and its Applications, 2009. CSA '09. 2nd International Conference* on (pp. 1-6). IEEE.

Goldstein, J., Mittal, V., Carbonell, J. and Kantrowitz, M., 2000, April. Multi-document summarization by sentence extraction. *In Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization* (pp. 40-48). Association for Computational Linguistics.

Gong, W., 2018. Regression Techniques Used in Hydrometeorology. Handbook of Hydrometeorological Ensemble Forecasting, pp.1-27.

Greenbacker, C.F., 2011, June. Towards a framework for abstractive summarization of multimodal documents. *In Proceedings of the ACL 2011 Student Session* (pp. 75-80). Association for Computational Linguistics.

Gupta, P., Pagliardini, M. and Jaggi, M., 2019. Better word embeddings by disentangling contextual n-gram information. arXiv preprint arXiv:1904.05033.

Gurney, K., 2014. *An introduction to neural networks*. 2nd edition. CRC press.

Hanczar, B. and Nadif, M., 2018, September. Controlling and Visualizing the Precision-Recall Tradeoff for External Performance Indices. *In Joint European Conference on Machine* Learning and Knowledge Discovery in Databases (pp. 687-702). Springer, Cham.

Harwell, M.R. 2011. Research design in qualitative/quantitative/mixed methods. 2nd ed. C.F. CONRAD & Ro. C. SERLIN, eds. Thousand Oaks, CA: SAGE Publications.

Haytam, Z., 2019. Implementing a flexible neural network with backpropagation from scratch | Awaiting Bits. [online] Available at: https://blog.zhaytam.com/2018/08/15/implement-neural-network-backpropagation/

Hirao, T., Nishino, M., Yoshida, Y., Suzuki, J., Yasuda, N. and Nagata, M., 2015. Summarizing a document by trimming the discourse tree. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 23(11), pp.2081-2092.

Hou, L., Hu, P. and Bei, C., 2017, November. *Abstractive Document Summarization via Neural Model with Joint Attention. In National CCF Conference on Natural Language Processing and Chinese Computing* (pp. 329-338). Springer, Cham.

Hsu, W.T., Lin, C.K., Lee, M.Y., Min, K., Tang, J. and Sun, M., 2018. A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss.

Huang, L., He, Y., Wei, F. and Li, W., 2010, April. Modeling document summarization as multi-objective optimization. *In 2010 Third International Symposium on Intelligent Information Technology and Security Informatics* (pp. 382-386). IEEE.

Jhalani, R., & Meena, Y.K, 2017. An abstractive approach for text summarization. *International Journal of Advanced Computational Engineering and Networking* (Vol. 3, pp.5–10).

Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp.255-260.

Kang, E., 2017. Hidden Markov Model. [online] Medium. Available at: https://medium.com/@kangeugine/hidden-markov-model-7681c22f5b9

Kasture, N.R., Yargal, N., Singh, N.N., Kulkarni, N. and Mathur, V., 2014. A survey on methods of abstractive text summarization. Int. J. Res. Merg. Sci. Technol, 1(6), pp.53-57.

Khan, A. and Salim, N, 2014. A review on abstractive summarization methods. Journal of Theoretical and Applied Information Technology, 59(1), pp.64-72.

Krauss, S.E., 2005. Research paradigms and meaning making: A primer. The qualitative report, 10(4), pp.758-770.

Krenker, A., Bešter, J. and Kos, A., 2011. Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, pp.1-18.

Kriesel, D., 2007. A brief introduction on neural networks.

Kyoomarsi, F., Khosravi, H., Eslami, E., Dehkordy, P.K. and Tajoddin, A., 2008, May. Optimizing text summarization based on fuzzy logic. In *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on* (pp. 347-352). IEEE.

Landthaler, J., Waltl, B., Huth, D., Braun, D., Stocker, C., Geiger, T. and Matthes, F., 2017. Extending Thesauri Using Word Embeddings and the Intersection Method. In ASAIL@ ICAIL.

Le, Q. and Mikolov, T., 2014, January. Distributed representations of sentences and documents. *In International Conference on Machine Learning* (pp. 1188-1196).

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), p.436.

Lin, C.Y., 2004. Rouge: A package for automatic evaluation of summaries. Text Summarization Branches Out.

Lipton, Z.C., Berkowitz, J. and Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

Logeswaran, L., Lee, H. and Radev, D., 2016. Sentence ordering using recurrent neural networks.

Lu, C.J. and Wang, Y.W., 2010. Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, *128*(2), pp.603-613.

Luhn, H.P., 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), pp.159-165.

Malhotra, R., 2015. A systematic review of machine learning techniques for software fault prediction. Applied Soft Computing, 27, pp.504-518

Mallett, D., Elding, J. and Nascimento, M.A., 2004, April. Information-content based sentence extraction for text summarization. *In Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on* (Vol. 2, pp. 214-218). IEEE.

Marsland, 2015. *Machine Learning an Algorithmic Perspective*. 2nd edition. New York. Chapman & Hall/ CRC.

McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E., 2006. A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. AI magazine, 27(4), p.12.

Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013a. Efficient estimation of word representations in vector space.

Mikolov, T., Karafiát, M., Burget, L., Černocký, J. and Khudanpur, S., 2010. Recurrent neural network-based language model. *In Eleventh Annual Conference of the International Speech Communication Association*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013b. Distributed representations of words and phrases and their compositionality. *In Advances in Neural Information Processing Systems* (pp. 3111-3119).

Mittal, N., Agarwal, B., Mantri, H., Goyal, R.K. and Jain, M.K., 2014. Extractive Text Summarization.

Mlitwa, N.W., Jp, V. & Belle. 2011. Integration of e-learning systems into academic programmes in South African universities.

Moawad, I.F. and Aref, M., 2012, November. Semantic graph reduction approach for abstractive text summarization. *In Computer Engineering & Systems (ICCES), 2012 Seventh International Conference on* (pp. 132-138). IEEE.

Mohammed, M., Khan, M.B. and Bashier, E.B.M., 2016. *Machine learning: algorithms and applications*. Crc Press.

Moratanch, N. and Chitrakala, S., 2016, March. A survey on abstractive text summarization. *In Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on* (pp. 1-7). IEEE.

Myers, M.D., 2013. *Qualitative research in business and management*. Sage.

Nallapati, R., Xiang, B. and Zhou, B., 2016. Sequence-to-sequence rnns for text summarization.

Nazeer, K.A. and Sebastian, M.P., 2009, July. Improving the Accuracy and Efficiency of the K-means Clustering Algorithm. *In Proceedings of the World Congress on Engineering* (Vol. 1, pp. 1-3).

Neuman, W.L., 2013. *Social research methods: Qualitative and quantitative approaches*. Pearson education.

Nilsson, N.J., 2014. *Principles of artificial intelligence*. Morgan Kaufmann.

Oates, B.J. 2005. *Researching information systems and computing*. Sage.

Okoli, C. 2017. A Guide to Conducting a Standalone Systematic to Cite this Version: A Guide to Conducting a Standalone Systematic Literature Review.

Okoli, C., & Schabram, K. 2010. A Guide to Conducting a Systematic Literature Review of Information Systems Research. SSRN Electronic Journal, 10(26), 1–51. https://doi.org/10.2139/ssrn.1954824

Over, P., Dang, H. and Harman, D., 2007. DUC in context. Information Processing & Management, 43(6), pp.1506-1520.

Oya, T., Mehdad, Y., Carenini, G. and Ng, R., 2014. A template-based abstractive meeting summarization: Leveraging summary and source text relationships. In Proceedings of the 8th International Natural Language Generation Conference (INLG) (pp. 45-53).

Paulus, R., Xiong, C. and Socher, R., 2017. A deep reinforced model for abstractive summarization.

Ragunath, R. and Sivaranjani, N., 2015. Ontology Based Text Document Summarization System Using Concept Terms. ARPN Journal of Engineering and Applied Sciences, 10 (6), pp.2638-2642.

Ramezani, M. and Feizi-Derakhshi, M.R., 2015. Ontology-Based Automatic Text Summarization Using FarsNet. Advances in Computer Science: An International Journal, 4(2), pp.88-96.

Reiter, E. and Dale, R., 1997. Building applied natural language generation systems. Natural Language Engineering, 3(1), pp.57-87.

Rokach, L. and Maimon, O., 2005. Clustering methods. In Data Mining and Knowledge Discovery Handbook (pp. 321-352). Springer, Boston, MA.

Rossiello, G., Basile, P., Semeraro, G., Di Ciano, M. and Grasso, G., 2016. Improving Neural Abstractive Text Summarization with Prior Knowledge.

Rudovic, O., Lee, J., Dai, M., Schuller, B. and Picard, R., 2018. Personalized Machine Learning for Robot Perception of Affect and Engagement in Autism Therapy.

Rush, A.M., Chopra, S. and Weston, J., 2015. A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685.

Russell, S., Norvig, P. and Intelligence, A., 1995. A modern approach. Artificial Intelligence. Prentice-Hall, Englewood Cliffs, 25(27), pp.79-80.

Saunders, M., Lewis, P. & Thornhill, A., 2012. Research methods for business students. 6th edition, Pearson Education Limited.

Saunders, M. N., Saunders, M., Lewis, P. and Thornhill, A. (2009), Research methods for business students, 5th ed, Pearson Education Limited, England.

Schank, R.C., 1972. Conceptual dependency: A theory of natural language understanding. Cognitive psychology, 3(4), pp.552-631.

Schuld, M., Sinayskiy, I. and Petruccione, F., 2015. An introduction to quantum machine learning. Contemporary Physics, 56(2), pp.172-185.

See, A., Liu, P.J. and Manning, C.D., 2017. Get to the point: Summarization with pointer-generator networks. arXiv preprint arXiv:1704.04368.

Shalev-Schwartz,S., Ben-David,S., 2014. *Understanding Machine Learning*. Cambridge University Press.

Silverman, D., 2013. *Doing qualitative research: A practical handbook*. London: Sage.

Smyth, R., 2004. Exploring the usefulness of a conceptual framework as a research tool: a researcher's reflections. Issues in Educational Research, 14(2), p.167.

Sridharan, R., 2014. Gaussian mixture models and the EM algorithm.

Sundermeyer, M., Oparin, I., Gauvain, J.L., Freiberg, B., Schlüter, R. and Ney, H., 2013, May. Comparison of feedforward and recurrent neural network language models. *In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 8430-8434). IEEE.

Sundermeyer, M., Schlüter, R. and Ney, H., 2012. LSTM neural networks for language modeling. *In Thirteenth Annual Conference of the International Speech Communication Association*.

Sunitha, C., Jaya, A. and Ganesh, A., 2016. A Study on Abstractive Summarization Techniques. *In Indian Languages. Procedia Computer Science*, 87, pp.25-31.

Sutskever, I., Jozefowicz, R., Gregor, K., Rezende, D., Lillicrap, T. and Vinyals, O., 2015. Towards principled unsupervised learning.

Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. *In Advances in neural information processing systems* (pp. 3104-3112).

Tantithamthavorn, C., McIntosh, S., Hassan, A.E. and Matsumoto, K., 2016, May. Automated parameter optimization of classification techniques for defect prediction models. *In 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)* (pp. 321-332). IEEE.

ThippeSwamy , R.D.K., 2017. Training Feed forward Neural Network with Back Propagation Algorithm. *International Journal of Engineering and Computer Science*, 6(1).

Wahyuni, D. 2012. The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies. JAMAR: *Journal of Applied Management Account Research*, 10(1): 69–80.

Wang, L., 2016. Summarization and Sentiment Analysis for Understanding Socially Generated Content (Doctoral dissertation, Cornell University).

Wang, L., Raghavan, H., Castelli, V., Florian, R. and Cardie, C., 2016. A sentence compression-based framework to query focused multi-document summarization.

Wang, S., Zhao, X., Li, B., Ge, B. and Tang, D., 2017, June. Integrating extractive and abstractive models for long text summarization. *In Big Data (BigData Congress), 2017 IEEE International Congress on* (pp. 305-312). IEEE.

Warden, S. C., 2011. Reader: Case study research - methodology and design. Btech Public Relations, Research Methodology, Cape Peninsula University of Technology, Cape Town.

Wiles, R., Crow, G., & Pain, H., 2011. Innovation in qualitative research methods: a narrative review. Qualitative Research, 11(5), pp.587-604.

Williams, C. 2011. Research methods. Journal of Business & Economics Research (JBER), 5(3).

Wilson, J., 2010. Essentials of business research: a guide to doing your research project. SAGE Publications, p.7.

Yin, R. K., 2013. Case study research: Design and methods, Sage publications.

Zaremba, W., Sutskever, I. and Vinyals, O., 2014. Recurrent neural network regularization.

Zell, A., 1994. Simulation Neuronaler Netze [Simulation of Neural Networks] (in German) (1st ed.). Addison-Wesley. p. 73.

Zeng, X. and Zhang, Y., 2013. Development of Recurrent Neural Network Considering Temporal Spatial Input Dynamics for Freeway Travel Time Modeling. Computer Aided Civil and Infrastructure Engineering, 28(5), pp.359-371.

Zhou, Q., Yang, N., Wei, F. and Zhou, M., 2017. Selective encoding for abstractive sentence summarization.

Zoubin, G. 2004. Unsupervised Learning. 1st ed. Gatsby Computational Neuroscience Unit University College London, UK.

APPENDICES

APPENDIX A: Document details

Documents	Words	Sentences	Paragraphs
1	4512	212	183
2	3432	179	170
3	4322	205	188
4	4812	262	191
5	8101	411	391
6	3931	187	179
7	4596	226	202
8	4620	226	221
9	4637	253	230
10	6867	326	281
11	5351	230	216
12	3820	205	193
13	4915	231	211
14	11303	603	510
15	9319	412	381
16	8095	378	346
17	3394	174	174
18	4590	212	214
19	6316	310	286
20	4495	220	204
21	3234	163	157
22	7259	356	330
23	4248	188	183
24	4486	249	244
25	3465	167	161
26	4063	206	167
25	3102	161	161
26	9113	500	458
29	9355	487	413
30	5167	234	221
31	6306	263	247

32	6935	333	270
33	10079	483	424
34	6521	314	278
35	5415	256	237
36	3354	153	152
37	3196	174	161
38	3895	238	196
39	5972	293	281
40	10056	460	436
41	3229	162	156
42	10189	559	475
43	6004	329	338
44	8374	449	446
45	8051	400	352
46	3237	158	155
47	14299	655	605
48	3749	206	157
49	3516	169	157
50	6160	283	265

Google drive link to access the program:
https://drive.google.com/drive/folders/1Lja3Wzc1UKOSQQ-P-NMN3yTM5SfckDiC

APPENDIX B: Results of ROUGE scores

			ROUGE-1			ROUGE-2			ROUGE-3			ROUGE-L			ROUGE -W		
Documents	Run Time	Number of sentences generated	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
1	54,07	4	42,11	21,82	28,74	12,5	6,42	8,48	1,82	0,93	1,23	45,23	26,15	33,14	19,57	5,61	8,72
2	61,02	3	36.84	12.39	18,54	8,11	2,68	4,03	2,78	0,9	1,36	43,51	17,55	25,01	24,54	4,54	7,66
3	57,6	3	40	20,95	27,5	7,41	3,85	5,06	1,89	0,97	1,28	41,24	24,06	30,39	23,14	7	10,74
4	65,24	3	37,5	18,75	25	10,91	5,41	7,23	1,85	0,91	1,22	31,5	17,68	22,65	15,32	3,9	6,22
5	54,57	3	40	14,29	21,05	5,13	1,8	2,67	0	0	0	39,2	16,62	23,34	17,87	3,62	6,01
6	54,03	3	26,56	16,35	20,24	1,59	0,97	1,2	0	0	0	28,18	18,8	22,56	12,65	4,03	6,11
7	67,25	4	47,92	21,5	29,68	8,51	3,77	5,23	0	0	0	40,03	20,52	27,14	22,29	5,09	829
8	48,43	3	33,33	24,49	28,24	8,45	6,19	7,14	0	0	0	34,39	26,6	30	18,3	7,34	10,47
9	57,99	3	28,57	14,14	18,92	2,08	1,02	1,37	0	0	0	30,96	17,23	22,14	16,85	4,63	7,27
10	66,61	3	45,95	16,35	24,11	0	0	0	0	0	0	47,12	19,92	28	26,67	5,5	9,12
11	49,42	2	33,85	20	25,14	1,56	0,92	1,16	0	0	0	32,7	21,1	25,65	16,78	5,49	8,28
12	65,22	3	28,33	15,32	19,88	3,39	1,82	2,37	0	0	0	29,74	17,81	22,28	15,3	4,37	6,79
13	57,43	3	54,72	28,43	37,42	15,38	7,92	10,46	5,88	3	3,97	42,53	24,65	31,21	23,48	6,77	10,5
14	50,84	3	28,57	15,69	20,25	1,82	0,99	1,28	0	0	0	23,8	14,44	17,97	12,47	3,66	5,66

15	70,55	4	30	12	17,14	0	0	0	0	0	0	28,85	13,44	18,34	15,6	3,4	5,58
16	53,13	2	27,27	12,5	17,14	2,33	1,05	1,45	0	0	0	31,5	16,44	21,61	17,52	4,36	6,98
17	45,98	3	52,38	21,36	30,34	7,32	2,94	4,2	5	1,98	2,84	47,06	22,28	30,25	27,57	6,11	10,01
18	53,97	2	39,13	15,52	22,22	11,11	4,35	6,25	2,27	0,88	1,27	39,3	18,18	24,86	22,63	4,81	7,93
19	54,65	2	48,08	21,74	29,94	9,8	4,39	6,06	4	1,77	2,45	35,49	18,32	24,16	21,42	4,94	8,03
20	50,22	3	35,14	24,76	29,05	4,11	2,88	3,39	1,39	0,97	1,14	37,76	28,21	32,3	20,19	8,44	11,91
21	56,42	3	31,25	21,05	25,16	4,76	3,19	3,82	0	0	0	29,85	21,48	24,98	15,66	5,52	8,16
22	64,56	4	25,64	9,09	13,42	2,63	0,92	1,36	0	0	0	29,47	12,42	17,47	16	2,98	5,02
23	58,78	2	51,72	28,3	36,59	12,28	6,67	8,64	1,79	0,96	1,25	44,58	26,97	33,61	20,84	6,31	9,69
24	58,49	2	61,02	34,95	44,44	31,03	17,65	22,5	12,28	6,93	8,86	58,49	36,77	45,15	34,46	10,81	16,46
25	50,81	3	34,72	24,51	28,74	7,04	4,95	5,81	1,43	1	1,18	30,03	22,47	25,7	14,75	5,62	8,14
26	57,94	3	37,21	13,68	20	4,76	1,72	2,53	0	0	0	32,11	13,94	19,44	16,24	3,49	5,75
27	46,44	3	28,07	14,16	18,82	0	0	0	0	0	0	29,18	16,5	21,08	15,16	4,75	7,24
28	58,13	3	59,38	17,43	26,95	12,9	3,7	5,76	6,67	1,87	2,92	59,03	21,26	31,26	34,8	5,79	9,93
29	55,07	2	43,28	27,88	33,92	10,61	6,8	8,28	6,15	3,92	4,79	38,03	26,36	31,14	21,22	8,41	12,05
30	61,21	3	47,54	24,79	32,58	8,33	4,31	5,68	0	0	0	49,13	28,55	36,12	27,21	8,09	12,47
31	66,26	3	35,82	21,62	26,97	10,61	6,36	7,95	1,54	0,92	1,15	34,99	22,97	27,73	18,82	6,66	9,84
32	56,76	4	47,73	17,65	25,77	11,63	4,24	6,21	2,38	0,85	1,26	45,27	19,76	27,51	25,66	5,49	9,04
33	46,27	3	24,66	16,98	20,11	0	0	0	0	0	0	26,75	19,6	22,63	13,3	5,44	7,72

34	57,5	2	33,33	15,15	20,83	6,82	3,06	4,23	4,65	2,06	2,86	33,24	17,23	22,7	19,89	4,93	7,9
35	61,11	2	48,15	12,26	19,55	0	0	0	0	0	0	47,32	15,14	22,94	27,93	4,36	7,54
36	52,13	2	48,98	20,69	29,09	8,33	3,48	4,91	2,13	0,88	1,24	47,39	23,11	31,07	27,05	6,8	10,86
37	57,97	3	21,88	6,6	10,14	0	0	0	0	0	0	24,78	9,13	13,35	13,91	2,47	4,2
38	66,84	2	50	23,15	31,65	18,37	8,41	11,54	10,42	4,72	6,49	50,45	26,56	34,8	30,42	8,68	13,51
39	53,45	4	31,82	13,33	18,79	0	0	0	0	0	0	26,65	12,91	17,39	14,18	3,2	5,23
40	46,28	3	35,19	19	24,68	7,55	4,04	5,26	1,92	1,02	1,33	36,29	21,72	27,17	20,18	5,96	9,2
41	60,99	3	48,98	21,24	29,63	4,17	1,79	2,5	0	0	0	41,39	20,63	27,53	21,95	5,22	8,44
42	49,89	4	36,11	12,04	18,06	2,86	0,93	1,41	0	0	0	37,23	14,9	21,29	20,49	3,94	6,61
43	61,23	3	43,08	26,92	33,14	14,06	8,74	10,78	1,59	0,98	1,21	43,59	29,47	35,16	24,36	8,52	12,62
44	53,85	3	40,38	18,42	25,3	11,76	5,31	7,32	4	1,79	2,47	41,31	21,48	28,26	23,27	5,65	9,09
45	87,71	3	46,88	28,57	35,5	7,94	4,81	5,99	1,61	0,97	1,21	44,16	29,23	35,18	25,17	8,15	12,31
46	63	3	26	11,61	16,05	6,12	2,7	3,75	4,17	1,82	2,53	30,44	15,55	20,58	17,26	4,23	6,79
47	57,52	3	48	23,76	31,79	20,41	10	13,42	10,42	5,05	6,8	48,53	27,01	34,71	27,21	7,65	11,94
48	55,68	3	32,88	23,3	27,27	6,94	4,9	5,75	1,41	0,99	1,16	38,19	28,67	32,75	19,11	7,85	11,13
49	67,29	4	38,71	23,08	28,92	6,56	3,88	4,88	1,67	0,98	1,23	40,57	26,36	31,96	19,25	6,09	9,26
50	55,56	3	46,77	25,66	33,14	19,67	10,71	13,87	13,33	7,21	9,36	43,76	26,54	33,04	26,42	7,94	12,21