Machine Learning in Text Summarization

DR SOM GUPTA
PGT CS, KVS, India

Abstract— Automatic text summarization, an approach to generate summaries for informal text helps in saving time during information retrieval and other related tasks. Although various kinds of techniques like fuzzy logic, and other soft computing skills, NLP have been widely used to achieve this result, the paper analyses the research done in this field using supervised, unsupervised and reinforcement learning. The paper also analyses the works done using deep learning approaches for both the extractive and abstractive summarization. The paper compares the evaluation results obtained by various approaches for a particular dataset, discusses about the popular datasets used for the research purpose; pros and cons of the various approaches, open challenges and the future directions in this field of research, The paper cites the famous research works from sources like IEEE, Springer, ACL Anthology, ACM libraries to do the analysis. The paper serves as the beginning point for the novel researchers who wish to apply ML based approaches for the text summarization purpose.

Index Terms- Machine Learning; Deep Learning; Text summarization; CNN; RNN; SVM; GRU

I. INTRODUCTION

The data in the digital world is increasing exponentially and the easy access to the Internet and the availability of many tools and the content related websites, has raised the need of reducing the text in order to retrieve the desired result quickly. Text summarization is to condense the data in such a manner that the essence of the documents remains preserved. The aim of automatic text summarization is to generate summaries which are fluent, coherent, concise, and non-redundant. Even when we search on google, the important sentences are displayed as summaries. Text summarization has been widely used in search engines, question-answering sites and recommendation engines. Few researchers have mentioned the following four as the parameters to decide the efficiency of a summary: Information Coverage, Information significance, Information Redundancy and Text cohesion.

The first work in this field is reported from 1958 with the research of Luhn [1] where they used the frequency to find the important sentences. From there onwards, a lot of research started taking place where the various techniques like graph-based, statistical, linguistic, machine-learning, deep learning and soft computing techniques like genetic programming, fuzzy logic, optimization techniques like swarm optimization, greedy approaches, linear programming, etc were used to generate the automatic summaries. Figure 1 displays the relation of Machine learning with the Deep Learning approaches. The first research using the machine learning techniques was reported in 1969 with the work by Edmundson [2] where they utilised the features like position, frequency, cue words and document structure to extract the important sentences.

According to how the summaries are generated, the summaries are either extractive or abstractive. Extractive summaries are those where the summary is generated by arranging the important extracted sentences by extracting the important key phrases whereas the abstractive summaries are those which contain the novel sentences by either the substitution, deletion or editing of the phrases. Abstractive summaries are closer to human-written summaries and look more meaningful. Figure 2 gives the picture of overall flow of extractive summarization techniques whereas Figure 3 gives the overall flow of abstractive summarization techniques. Abstractive summaries help overcome the grammatical mistakes of the sentences. These techniques use the bottom up approach to generate the summaries. Sentence extraction and the Feature Score calculation are the common techniques to create the extractive summaries. Sentence Compression, Paraphrasing, Reordering, Generalising are few techniques which are used as part of the abstractive summarization approach. Encoder-Decoder architectures along with the Natural Language Generation are famous for creating abstract summaries. The first work on abstractive summarization using deep learning models was reported in 2015 which was based upon the encoder decoder model.

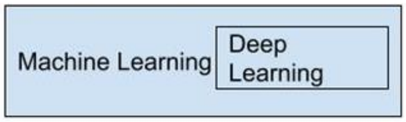


Figure 1: Relationship between Machine Learning and Deep Learning

For text summarization, there are many models which have been successfully used by various researchers. Natural Language Processing, Topic Based Information Retrieval Models, Machine Learning Models and Deep Learning Models are the mainly used models. As many times, it is not only one topic which determines the importance of a sentence, it is important to include various indicators and combine them to find the important sentences. These techniques consider the summarization task as a classification problem which intends to find whether the sentence should be included in the summary or not. Nowadays data-driven approach is one of the famous approaches, to find the important sentences and is thus the reason that the research is more towards using machine learning, neural networks , or the deep learning models. Also these models help reduce the dependency to the linguistic analysis tools and the other preprocessing tools.

Machine Learning approaches can be divided into: Supervised and Unsupervised approaches. Supervised approaches involve the usage of labelled training data whereas Unsupervised approaches do not require the labelled training data.

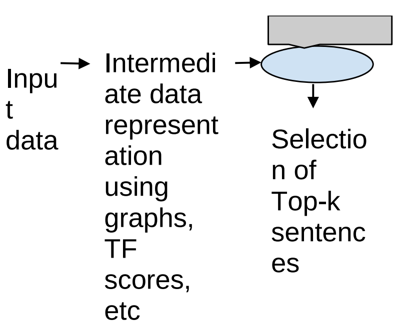


Figure 2: Overall flow of extractive summarization techniques

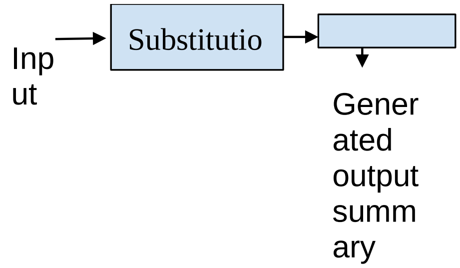


Figure 3: Overall flow of abstractive summarization techniques

In the supervised learning approach, the problem of finding the important sentences is considered as the binary classification problem. Classifiers are the base of these approaches. In general, in this approach a supervising algorithm is used for training the summarizer to find the important segments and mostly the feature vectors are used for this purpose over which the approaches are applied. For generating the feature vectors, most of the researchers emphasise on using the domain-independent features and using the principles of generality and the subjectivity for choosing the features. Which approach is used determines the efficiency and accuracy of the system. Preparing the labelled annotated data for the training is a time- consuming task. Annotator agreement is also one parameter and it is difficult to come to the conclusion which sentence is important for the summary as many times different annotators select different sentences for inclusion to the summary [3]. In a paper by Nenkova [3], they mentioned that for the multi-document summarization supervised approaches have given better results than the unsupervised approaches. Figure 4 gives the overall picture of how the machine learning algorithms work.

Machine learning algorithms capture the high dimensional features very nicely and also the non linear relationships. This helps understand the dynamic data more nicely. Narayana et al. [4] used the maximum likelihood cross entropy loss along with the policy gradient to improve the evaluation results. CNN and DailyMailset were used for the evaluation of results.

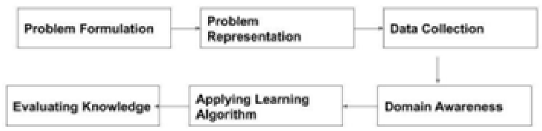


Figure 4: General Flow of Machine Learning Based Approaches

Text summarization has been successfully used in almost all the domains. Summarization for Movie Reviews, Product Reviews, Emails, Software Artifacts, Scientific Literatures, Books, News, Weather Forecasts, Emails, Stock Market are some famous artifacts where the researchers have widely used this technique [5].

The research paper reviews the usage of Machine Learning Approaches for the text summarization purpose. The main focus is to find out the various Machine Learning approaches being used for automatic summarization. The paper investigates the answers to the following questions:

Q1: What are the main Machine Learning Approaches used for this task

Q2: What are the main tasks during summarization process where the ML techniques are used

Q3: How the results vary with the application of Supervised and Unsupervised approaches

Q4: What are the strengths and the limitations of using ML techniques for summarization

Q5: What is the status of Deep Learning for Text Summarization

Q6: How the results vary with the usage of Deep Learning than by simple ML Based Approaches

Q7: What are the challenges associated with using ML Techniques

The paper is organised as follows: Section 2 states the famous Machine Learning Techniques being used for summarization. Section 3 gives the overview of research done using Machine Learning Techniques for generating extractive and abstractive summaries.

Section 4 describes the status of deep learning techniques in text summarization. Section 5 describes the evaluation metrics being used when either the Machine Learning Techniques are used or Deep Learning Based Approaches. Section 6 describes the methodology used for the survey and the statistics about research in this field. Section 7 discusses the strengths and the limitations of using Machine Learning in Text Summarization. Section 8 discusses the future directions and then finally the Conclusion.

II. RESEARCH METHODOLOGY

We have used the Systematic Literature Review for the conduct of Review of Works in the field of Text Summarization involving the use of Machine Learning Techniques. We restrict our work to pure Machine Learning Approaches and exclude the deep analysis of Neural Networks and the Deep Learning Models, as they demand another work in this field. We have already included the Deep Learning Models application on Text Summarization including both the Extractive and Abstractive Techniques in our other work [6].

While we were working on the summarization techniques, we found that they involve the use of Natural Language Processing, Information Retrieval Techniques, Machine Learning Techniques and now extended to Deep Learning also. We found that there are so many research works which classify their works on the basis of many factors like Type of Summary, Type of Technique, Type of Application; but there is a lack of depth classification according to the technique. For most of the works, there is hardly one paragraph stating the state of Machine Learning in Text Summarization. Thus we decided to conduct one exclusive work for this area only. And we wish to include all the background information including the exhaustive mentioning of the various works, different techniques available, strengths and limitations and other inclusion of other aspects of the research and analysis. Figure 4 gives an overview of the research technique that we used to do our research.

Our Approach included the selection of Topic, followed by selection of libraries, followed by reading the Titles and the Abstracts to understand its fitness to our work. We have selected some Keywords and phrases on the basis of which we performed our selection.

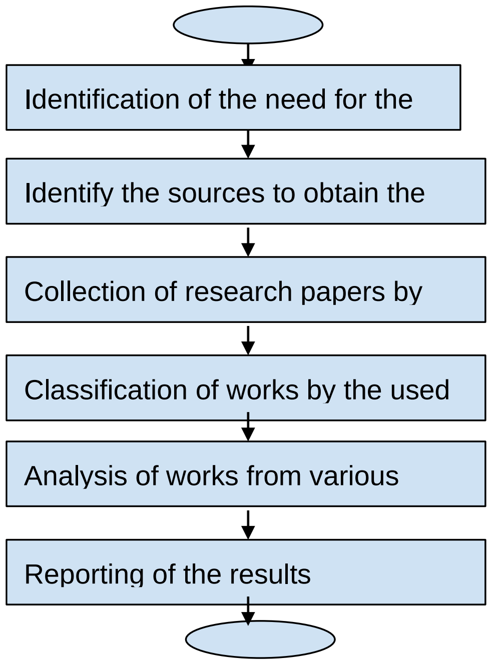


Fig 4: Systematic Literature Review

KEYWORDS USED:

- Machine Learning in Summarization
- Text Summarization
- Machine Learning in Text Summarization
- Machine Learning for Summarization

LIBRARIES USED

We have used the popular libraries:

- ACL Anthology
- Elsevier
- Springer
- IEEE Xplore
- ACM
- Google Scholar

Analysis from Various perspectives:

1. Identification of Journals for the mentioned area
2. Identification of various techniques used for the mentioned area
3. Identification of various datasets used for performing the experiments.
4. Identification of various features being used for performing the mentioned task
5. Identification of problems and challenges in using the mentioned approach
6. Identification of evaluation metrics for evaluating the techniques for the mentioned task
7. Identification of future research areas

III. RELATED WORK

The use of machine learning for text summarization started in 1965 when Edmundson [2] used the Cue Phrases, High Frequency Words, Sentence Location, and Title and Headings, for creating the vectors for the text and then used Naive Bayes for training the text to find the important sentences to be selected for summary generation.

In 1997, Lin et al. [7] used the Decision Trees to find the important topics by considering the positional properties for creating the abstracts from the text.

In 1999, Aone [8] proposed a model known as DimSum and they also used the Bayesian model incorporating the positional information of the text and inverse frequency term with the term frequency in addition to the Edmundson model [2].

Chuang et al. [9] in 2000 created an approach to automatic Text Summarization that generates the summary by extracting the sentence segments. The sentence segments were represented by feature vectors. They used 23 features to create the vector. These 23 features were grouped into three classes: Non-Structural, Rhetorical, collective descriptions of Rhetorical features. Following are the features used by them: Paragraph Number, Offset in the paragraph, Number of Bonus Words, Number of Title Words, Term Frequency, Antithesis, Cause, Circumstances, Concession, Condition, Contrast, Detail, Elaboration, Example, Justification, Means, Otherwise, Purpose, Reason, Summary Relation, Weight of Nucleus, Weight of Satellite, and max level. After the creation of the feature vector, Decision Trees and Naive Bayes classifiers were used for the training purpose.

In 2001, Conroy et al [10] used the Hidden Markov Models for extracting the sentences. They used HMM models because these models capture the local dependencies better. Mureson et al. [11] in the same year used the decision trees C4.5 for the email summarization.

In 2002, Amini et al [12] used a semi-supervised learning approach to create the text. Summaries. In their model they used the amalgamation of unlabeled and labelled dataset.

In 2005, Amini et al. [13] used the logistic regression models along with the clustering approaches to rank the sentences with the minimum ranking loss for creating adaptable summaries. Osborne [14] used the maximum entropy model to extract important sentences from the document. The reason for choosing it is that they do not assume independence, and the utilization of external knowledge was possible using these models.

Hirao also [15] used the SVM approach to find the important sentences from the text. For their model to create the vectors with values between 0 and 1. They used the following features:- Position of Sentence , Length of Sentence, Weight of Sentence (TF-IDF), Density of Keywords, Named Entities, Conjunctions, Functional Words and Part of Speech, Semantic Depth of Nouns, Document genre, Symbols, Conversation, and Assertive Expressions. They found that SVM outperformed C4.5 and C5.0, decision trees and boosting approaches.

Neto et al. [16] in 2003 used Mean TF-ISF, Sentence Length, Sentence Position, Similarity To Title, similarity To Keywords, Sentence To Sentence Cohesion, Sentence To Centroid cohesion, Depth in tree, Indicator of Main Concepts, Occurrence of Proper Names, Occurrence of Non Essential Information, and then trained the model using C4.5 and Naive Bayes. In 2003, [17] Fattah used Naive Bayes, Entropy Models and SVM for creating the summaries for Multi-Document Summarization. DUC 2002 corpus was used for the research purpose by them.

Bollegalla et al. in 2005 [18] used Machine learning for the task of ordering sentences. They used the approach for multi document summarization. They used weighted Kendall Coefficient and Average Continuity for the evaluation of summaries. For ordering sentences, they used the concept of pairwise preference function. Suppose there are two sentences s1 and s2 out of total sentences SAll, they asked the annotators or experts to tell the precedence of s1 and s2. If the answer by an expert is more than 0.5, then that sentence gets the higher probability. Along with the preference, they considered publication date of sentence, unique identifiers in the sentences, and sentence position in the text. Other than preference and chronological features, probabilistic features and the topic relevance features were also used for creating the vector for training for the text by the system. Their approach helped in creating readable summaries.

Leite et al. [19] in 2008, used Flexible Naive Bayes, C4.5, SVM and Logistic Regression to summarize the Brazilian text and used the corpus TeMario. They observed that Naive Bayes and Logistic Regression outperformed C4.5 and SVM. Wong et al. in the same year used the Probabilistic SVM, Naive Bayesian and Co-Training to generate the extractive summaries.

In 2009, [20] used the SVM and Neural Networks to observe how they behave when applied to the task of text summarization. They found that both the models were able to capture the non-linear behaviour in the data. They used the following features for creating the vector:

Lexical chaining based, Sentence Length, Proper Nouns, Sentence Location,Sentence location,word frequency, relationship mapping, importance of topics, and complex network features like degree, minimal paths, locality index, matching index, k-cores, dilation and communiites.

Prasad et al. [21] in 2009 performed a study to observe how the machine learning algorithms behave for text summarization systems. They named their system Evolving Connectionist Text Summarizer. In their system they used the combination of machine learning and neural networks. For the machine learning model, they selected the features from the text. For selection of features, they used the Knowledge-poor and Knowledge-rich approaches concept. Knowledge-poor consisted of those approaches which try to assign the weights to the sentences based upon the features.

Whereas Knowledge-rich consisted of those which require domain knowledge.

In the same year, Kianmehr et al. [20] compared the SVM and Neural Networks and found that both the models perform almost the same. Moreover SVM takes less time and is less complex in comparison to Neural Networks. They also found that at a point, even increasing the number of layers in Neural Networks do not increase the efficiency and the results.

Raut et al. [22] in 2014, used Naive Bayes, SVM and Decision trees for summarization of online hotel reviews. They obtained 88 percent accuracy with the Naive Bayes algorithm, 83.5 percent with the SVM and 78.4 percent with the Decision trees. In a paper by Silva et al [23], mentioned the two ways of measuring compression rates for evaluating the extractive summaries. Horizontal Compression Rate is the summarization at sentence level. This is done by removing the unimportant and redundant words from the sentence. The horizontal rate compression is calculated by finding the ratio between the number of words in the original document to the number of words in summary.

Vertical Compression Rate is the number of sentences in the original document to the number of sentences in the summary. They used a feature-based approach along with the machine learning techniques to create the extractive summaries. They used 20 features which included all sentence based, word based and graph based.

Aggregate Similarity, Busy Path, Centrality, Heterogeneous Graph, TextRank were the graph based features used by them. Cue Phrase, Numerical Data, Position Paragraph, Title Resemblance, Sentence Length, Sentence Position in the Paragraph, Sentence Position in the Text were the sentence-level features used by them.

Proper Noun, Co-occurrence Bleu, Lexical Similarity, Co occurrence N Gram, TF-IDF, UpperCase and Word Frequency were used as the word-level features.

They removed the outliers and then used the Naive Bayes, MLP, SVM, KNN, AdaBoost and Random Forests as the classification approaches. For solving the problem of imbalancing with machine learning, the SMOTE [24] approach was used by them. They performed their experiments on a CNN dataset. According to their experiments Naive Bayes and AdaBoost performed the best.

Lanyo et al [25] in 2018, studied the application of machine learning in customer service platforms. They used Naive Baye and TextRank for these platforms. They observed that Naive Bayes gave better results in terms of accuracy and processing time. Day et al. [26] in 2018 created a statistical, machine learning and deep learning model to create the titles from the essay abstracts and create the summaries also. Zawbaa et al [27] in 2011 used SVM and the K-NN to summarize the Soccer Videos.

Kumar et al. [28] in 2018 used n-grams, presence of thematic words, presence of valid keywords, relative sentence length, sentence cohesion score and sentence position in the document to create the vector and then trained using Naive Bayes for Brown Corpus. Armouty et al in 2019 used the Support Vector Machines to find the important keywords from the Arabic News Documents.

Ahmed et al. [29] in 2019 also mentioned the use of machine learning techniques Clustering and Frequent Dataset for structured data summarization and the use of Naive Bayes, Decision Trees, Hidden Markov Model, Log Linear Models for unstructured data summarization.

Jo et al. [30] in 2019 proposed a text summarization tool based upon the machine learning algorithms for the news articles. The approach first converts the text into paragraphs and then the text is converted into the graphs where the edge represents the semantic relation between the words. Modified KNN approach is applied to create the summaries. Mohamed in the same year 2019, used the classification techniques and then evaluated the summaries using a modified version of F-Measure. In the same year Vivek et al. [31] used the Naive Bayes to train the Newspaper, Magazine, Blog Articles and Journal Articles.

Karmi et al. [32] in 2020 , created the 250 word summary on the reference spans. They observed how the citations help create the informative summaries. For this, they used the five machine learning algorithms SVM, Random Forests, Decision Trees, Multi-layer perceptron, AdaBoost and TF IDF to train the features. Anshuman Pattanaik in 2020 [33] applied the following algorithms for training the newspaper dataset to create the extractive summaries: KNN, Random Forests, Support Vector Machine, Decision Trees, Logistic Regression and Neural Networks. They only considered Sentence Length and TF-IDF as features to implement these algorithms. Even though they just used two features, they achieved an accuracy of 72 percent. This study shows the impact of just these two features in the field of summarization.

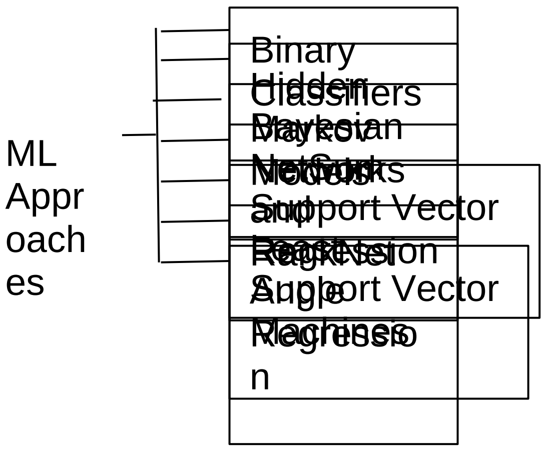


Figure 5: ML Approaches by Lloret et al.

Supervised Approaches in Text Summarization

The popularity of supervised approaches started with the work of Kupic et al. where they used the Naive Bayesian approach to categorise the sentence as important or not. Feature independence was assumed in their work and later more features were incorporated. They have been used for both the single document summarization as well as multi-document summarization.

1. Regression in Text Summarization: Here the aim is to fit the predicted scores to the target scores. Here labelling of the sentences is not more important. Support Vector Regression and Integer Linear Programming are two techniques which have been widely used for this task.

2. Classification in Text Summarization :
Decision Trees

They have been widely used for the inductive learning in the text summarization task. Many variations like ID3 and C4.5 have been used. These approaches are based upon the information gain. Mainly the instances are represented as the feature vectors. According to how the decision is made by the decision trees on the basis of attributes, the trees are classified into axis parallel and oblique. Axis parallel trees consider a single attribute for the split of decision trees whereas oblique trees consider the linear combination of features to split the trees. Lin et al. [41] , Knight et al. [42] also used the decision trees.

C4.5 [1]: It is a decision tree based classifier. Here the important features are extracted on the basis of information gain. The process is repeated till further the change in information gain does not occur. It is observed that this algorithm works very nicely on the generalisations.

Binary Classifiers: Markov Models Conroy et al. [10]: Hidden Markov Models are based upon the concept of joint probability distribution. The text is regarded as the graph of states. States are connected to each other. Transition Probability, Emission Probability, and Probability are used for indicating the relation between the states. States cannot be observed directly but the output depends upon the states.

Support Vector Machines: It is one of the powerful supervised approaches based upon the maximum-margin hyperplane where the focus is to separate the samples into two groups with the objective of maximum margin. Mandal et al. used the SVM, K-Nearest Neighbour and Decision Trees to generate the extractive summaries. It is a binary class problem. It is a classification based approach. When it is viewed from the point of view of summarization, the features are calculated from the sentences and the aim is to identify whether it is important or not. Thus the summarization task can be considered as a binary class problem. The main concept over which it works is the margin. It is non-probabilistic. It involves the identification of hyperplanes which divides the two classes. Distance between the margins are maximized and mean square error between the margins are minimized. It involves the use of kernels. Kernels help in processing the features internally. According to the type of value involved in the feature, the kernel is used. Few kernels which have been used are:

String Kernel for processing of the string values,

Spectrum Kernel for finding the mutual substrings,

S-S Kernel for finding the text classification,

L-W-S Kernel for matching subsequence to find the similarity between two text, and Context-S Kernel for sliding window concepts.

It was used by many researchers like Kandasamy et al. [31], Hirao et al. [15].

Pei et al. [64] also used the SVM to rank the sentences for the multi-document summarization.

Naive Bayesian[1]: It is a supervised approach and it is based upon the assumption that all the features are independent. The first studies on the usage of this technique for the summarization task was observed by the work of Kupic et al. [2] in 1995, where they used Naive Bayesian classifiers after finding the features Sentence Length, Fixed Phrase, Paragraph Feature, Thematic Feature and Uppercase feature to find the scores for the sentences. The scores for each sentence was estimated by calculating the conditional probability from the dataset. This approach with unigrams and bigrams was used by Khan et al. in 2020 to create the summaries for the movie reviews into the positive and negative and then over to it they used the graph based approach to generate the summaries. It is a classification-based algorithm and is a probabilistic classifier. It assumes the independence of features. It involves the calculation of conditional probabilities. Here the variance of each class is calculated.

$$P(s \subseteq S|F1, f2...fn) = \prod_{i=1}^{i=n} \frac{P(Fi|s)P(s)}{\prod_{i=1}^{i=n} P(Fi)}$$

Let the number of features be n.

s is the sentence from the total set S of the sentences of the text

P(Fi|s) and P(Fi) are calculated from the training data set.

Lanyo et al. [26], Neto et al. [16], Chuang et al. [9] used the Naive Bayesian approach for supervised learning in their approaches.

Maximum Entropy Model: It is a very powerful model. It is discriminative in nature. The aim is to maximize the entropy distribution while following the constraints. It is the measure of informative content. Entropy is maximum when the distribution is normal in nature.

Artificial Neural Networks: These are also based upon the machine learning techniques requiring the training dataset and the computation of error loss and others but for the computation rather than using the simple probability and other measures, use the graph like network architecture involving the input layer, hidden layer and the output layer. Many authors have used back propagation along with them to classify the features.

Most of the approaches utilizing neural networks follow the following approach:

Input text is fed to the system, the output is compared against the expected output, error is calculated from the output neuron, adjustment to the output is maintained and the process continues till the error change is minimal.

Conditional Random Fields: It is a statistical approach which is used to recognize patterns in machine learning. It is a discriminative probability based model based upon the concept of conditional probabilities. Here the text is considered as a graph. The vertices represent the random fields and are used for obtaining the conditional probabilities. Normalisation, and Weights and features are two main components of Conditional Random Fields.

Name of Approach	Used By	Dataset Used
Naive Bayesian	Neto et al. [16], Chuang et al. [9], Silva et al. [3]	Tipster [1], Engineering Data [2]

C4.5	Neto et al. [16], Chuang et al. [9], Mureson et al. [11]	Tipster [1]
Neural Networks		CNN, Wikipedia(svore)
Neural Networks	Sinha et al. [10]	DUC 2002
Convolutional Neural Networks	Zhang et al. [11]	DUC 2002, DUC 2004
SVM	Hirao et al. [15] , Silva et al. [23], Kianmehr et al.[48]	Text Summarization Challenge Dataset

Unsupervised Approaches in Text Summarization

Due to the fact that these approaches do not require much data corpora, they are getting popular. But as per the analysis done by various works, we have found that they are not able to achieve the results as par with the other techniques. Kohita et al. [63] used the Q-Learning with language models to perform the unsupervised summarization. They used the editorial agent to help perform the edit operations like Remove, Replace and Keep; whereas the language model decodes the sentence on the basis of action signals.

Ayodele et al. [16] used the unsupervised learning techniques to generate the summaries for incoming emails. They used the concept of inclusion of highly frequent words in the sentence and rearranging them to generate good summaries. They used the subject and the content field for the same.

1. Fuzzy Logic: Sheridan et al. used the fuzzy logic along with the evolutionary approaches
2. Cluster Based: K Nearest Neighbor is one of the most popular cluster based machine learning approaches which has been used in many researches. In a paper by Rahul et al. [21], have mentioned the works done by various researchers and have mentioned that atleast 14 percent of the works in this field used this approach.
3. Graph Based: Textrank is one of the algorithms which has been widely used in the research community to create the summaries. The approach is based upon calculating the similarity between the sentences in the document which is stored in terms of the Similarity Matrix. Here the sentences are represented in the form of vectors and then the similarity between the two vectors is either calculated by using the cosine similarity or any other powerful similarity measure. Similarity Matrix is represented in the form of a graph with vertices as the sentence and the edge as the similarity score.

Features Used for Summarization:

1. Statistical Features:

TF-ISF: Also known as Term Frequency and Inverse Sentence Frequency; is one of the features which has been used by almost all the researchers and works related to Summarization tasks.

Sentence Length: It helps get rid of irrelevant and very short sentences. For most of the research, normalized length is used.

Sentence Position: It is the position of sentence in the document. Earlier is the location of a sentence in the text, better is the chance of their inclusion to the summary.

Similarity to Title: Mainly the cosine measure is used to calculate the similarity.

Similarity to Keywords

Sentence To Sentence Cohesion: Let the sentence be S, and other sentences by S’. Similarity is calculated between the S and all S’ and then summed up. If the value comes closer to 1 that indicates the cohesiveness.

Sentence To Centroid Cohesion: Here rather than calculating the similarity of one sentence to all the sentences, first the centroid of the whole document is generated in terms of vector and then the sentences are compared with the centroid. If the value is closer to 1.0, that indicates that this sentence represents the central theme or is important for the document.

Paragraph Position

Format Based: Text in the bold, italics, underlined or having bigger font were extracted.

Numerical Data

Cue Phrases like “it is important”, “in the summary”, “in conclusion”, etc.

2. Linguistic Features[1]:
Main Entities

Proper Names: They refer to the place and the person, indicating their importance to the summary.

Occurence of Non Essential Information: They help filter out the irrelevant information which can be the cause of redundancy in the summary.

Occurrence of Anaphors: Anaphors refer to the pronouns and are those sentences which have been explained or referred to in the previous sentence. If these sentences are chosen to be the part of the summary, there is a chance of including the redundant sentences in the summary.

Lexical Similarity

Co-occurrence N Grams

3. Discourse Features:
They utilize the properties like whether the sentence has come just below the main headlines, etc.

4. Rhetorical Structure:
Like analyzing the causes, circumstances, antithesis, etc information to find the significance of the sentence fragments.

Kaikhah et al. [4] in 2004, used the neural networks to learn the characteristics of the text to create the text summaries. They used the training, feature fusion and the sentence selection in their approach. a 3 layered feedforward neural network was used for the training purpose. Conjugate gradient function was used for error function and the penalty function. Objective was to find the minima for energy function. To find the relationship between the sentences, feature fusion was used.

In 2005, Burges et al. [5] used the RankNet and In 2007, Score et al. used the RankNet approach to train the model for ranking the sentences. In 2010, the multiple perceptron layer architecture was used along with the fuzzy logic for generating the first layer input for the further genetic model.

In 2016, Singh et al. used the Restricted Boltzmann machine to generate the summaries.

Deep Learning for Text summarization

Deep Learning Models have emerged as a powerful machine learning tool. It is an extension of Neural Networks and Machine Learning Models. They attempt to imitate the human brain. Even though all are powerful themselves, because of the fact that the Deep Learning Models have the facility to extract the features also, they yield better results. Features are learnt at various levels and which makes it able to learn the various complex functions also [49]. With the smaller datasets, the machine learning algorithms work best but when the dataset is huge, deep learning models work better than the machine learning algorithms.

Neural networks are the models which consist of only one hidden layer of computation between the input and the output layer. When the Neural Networks contain more than 1 hidden layer, it becomes the deep learning model. Deep learning models are one of the wide applications of machine learning approaches. They have been used for generating both the extractive and abstractive summaries. Many times, the input data is so long that it becomes very challenging to retain the critical elements to create the summaries. As deep learning models have the ability to capture very complex non linear data, they can be used for these purposes. They help reduce the manual dependency in extracting features and having knowledge of linguistic rules. They also help capture multiple levels of representations. They learn the features automatically using various mechanisms like back propagation with the objective of minimizing the function. As the summarization process is highly dependent upon the features, thus deep learning models yield very good results for the summarization task. Not only for summarization tasks, but deep learning techniques have proven to be very effective for the NLP tasks.

Unsupervised approaches

Auto Encoder Decoder Model:

Azar et al. used the AEs and Restricted Boltzmann Machines to rank the sentences.

Seq2Seq is one of the popular models used for the summarization task which takes the sentences as the input and generates the sentences as the output. They map the variable length input text to the variable length output. Encoder and Decoder are the important techniques used in Seq2Seq models. The encoders take the input and encode them to a vector form or any relevant form like embeddings and a feedback which is also known as hidden layers is generated after every step. Word Embeddings is one of the popular ways to represent the sentence into the vector form as it helps in solving few of the serious problems in the field of machine learning like curse of dimensionality, and sparsity. Mainly Word2Vec, GloVe, FastText, and BERT are few popular ways to generate the word embeddings. Word2Vec consists of skip-gram and Continuous Bag of words to generate the embeddings. Encoders help capture the context of the text. Decoders are used to decode the target sentences. Most of the approaches which are built only using this model use the Beam Search Method.

Beam Search Models are for the prediction of output in the encoder decoder mechanism. In the Greedy approach, at each step the word with maximum conditional probability score is selected. Beam Search approach helps in finding the balance between the greedy approach and the exhaustive search mechanisms.

Shi et al.[26] used the Seq2Seq models and the beam search method for abstractive text summarization.

Attention Mechanisms:

The traditional models where the encoder-decoder architecture was followed, the problem of repetition, length, and exposure bias were there. In order to reduce these issues, the attention mechanism and other approaches in deep learning became popular.

The significant work using this mechanism was done by Rush et al. [44] in 2015, where the Bag of words model was used for generating the vector for the input text. A Deep Convolutional model was used at the encoder side. Max pooling layers were used along with the Convolutional layers. Attention mechanism was also used to capture the context in the summaries. Beam Search was used at the decoder side. Most of the researchers who used the encoder decoder model, have used this approach at the decoder side to get the appropriate results .

Supervised approaches:

Convolutional Neural Networks: They have been successfully used for image recognition and are now used for the summarization task also. They mainly use: weight sharing, pooling, multiple layers and local connections. In the convolutional model, the CNN layer uses the previous outputs from the CNN layer to find the local connections and then the pooling helps combine the semantically similar features into one.

Input data is represented in terms of vector spaces, and word embeddings by using methods like word2Vec. Flatten layers are used after the Convolutional layers to convert the 2 D matrices to the vector form. Fully connected layers are used to perform output processing. To minimize the overfitting, drop out along with the regularisation layer is used.

Zhang et al. [11] used the CNN layer followed by max pooling over the pre-trained vectors. They conducted the experiments for both the single document summarization and the multi-document summarization. CNN model was used for learning the sentence features, and ranking the sentences with the objective of minimising the cross entropy. The model was made in such a way that human interventions could be removed. Window size of 3 and the 400 filters were used for performing single document summarization which was increased to 600 filters for the multi-document summarization.

Recurrent Neural Networks: They are good at modelling the sequential data and help capture the syntactic and semantics from the word sequences.They have been widely used for the sequence prediction problems like word sequence prediction when one is writing, etc. Here the output is obtained not only from the existing data state but also from the previous outputs. In order to deal with the gradient explosion and vanishing effect, the below mentioned variations are used.

They can be encoder-decoder based RNN models or bidirectional RNN. In Bidirectional RNN, it consists of forward RNN and backward RNN. Forward RNN uses the input sequence from left to right whereas the Backward RNN reads the sequence from right to left. Here one representation is obtained by combining both the forward and backward RNN.

LSTM: Song et al. used the LSTM-CNN to construct the new sentences. In terms of summarization, these models are also known as sequence-to-sequence models. Anand et al. in 2022, used the LSTM based approach to generate the summaries for the legal documents. Convolutional and Recurrent Neural Networks were also used so as to consider the order of the input of data. As the RNN captured only short term information, LSTM was used where the sequence of data was long. They used input gate, memory gate, forget gate and output gate.CNN models along with the pooling were used before applying the LSTM to find the features from the input text.

GRU :

The Seq2Seq models suffer from repetition and capture of context. Gates are used to tackle these drawbacks. They help refine the representation of source text in the intermediate form.

These are variations of RNN models whose main function is to address the issue of exploding/vanishing gradients.Thus helps identify the long term dependencies in the text. They are the simplified form of LSTM and consist of reset and update gates. GRU takes less time than LSTM.

The research in the field of abstractive summarization was first reported by Rush et al. [44] in 2015 where they used the three models to perform the same: Bag of Words model, Convolutional Encoder, and attention-based mechanism. In the same year, Lopyrev [51] used the RNN models to generate the newspaper headline summaries.

Nallapati et al. [8] in 2016 used the Attentional Encoder-Decoder Recurrent Neural Network for text summarization. Their model used the bidirectional Encoder and unidirectional Decoder along with the soft attention. This work was further extended by Hasselqvist[7] in 2017 where they used the pointer generator model along with the query. Document and the query was fed as the input, sequence of words are passed to document encoder which is then passed to attentive decoder. For encoder and decoder, RNN with gates were used by them. The document is fed to the document encoder which uses the bidirectional RNN whereas the query is fed to the query encoder. Because of the fact that queries are shorter in length, they used unidirectional RNN. Decoder also uses unidirectional RNN along with the soft attention.

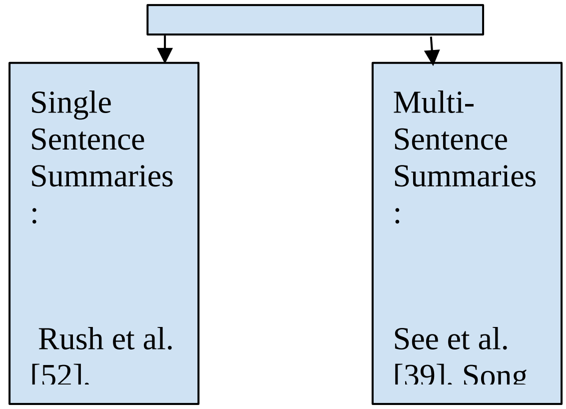
Lin et al. [41] used global encoding for text summarization. The Convolutional Gated Unit was used to perform global encoding.

Even though deep learning models are very powerful and help get good results, they require a lot of computation and a big training dataset is required. Deep learning models involve the usage of GPUs for performing the complex computations for the training purpose which limits the usage of deep learning models for the summarization purpose. Tuning of hyperparameters is also required for the deep learning approaches. If the deep learning models are not able to generalise properly for the large training dataset with a lot of parameters, it leads to errors during the testing phase.

Reinforcement Learning:

It is a type of Machine Learning technique where the system interacts with the environment to maximize the reward. The Sequential Markov Model is one example of Reinforcement Learning. At a particular timestamp, the agent includes the document and the previous extracted information. From this information, the agent will decide whether the sentence has to be included or not. Based on the decision, the agent will receive the reward on the basis of how good the decision was made. Final reward obtained by the agent determines the overall quality of the decision made by the system. These approaches have been mainly used for the controlling systems. They can be used to optimise the score functions for extractive summarization. The approach helps in reducing the need for hand crafted features. The first work in summarization using this technique was reported in 2012 with the work of Ryang and Abekawa where this process of extracting important sentences was considered as a search problem. Narayan et al. [69] used reinforcement learning for the task of sentence ranking for the generation of extractive summaries. They modelled the problem in terms of Agent which interacts with the environment. Environment consists of documents. Initially, the agent is randomly initialised. Then the agent reads the documents, predicts the relevance score for each sentence using policy which then uses the probability scores for the calculation. The agent is given a reward for finding how well the extracted summary resembles the golden-set summary. For reward, the mean of F Scores of ROUGE-1, ROUGE-2 and ROUGE-L were used. ROUGE-1 and ROUGE-2 helped capture the informativeness due the fact that they consider the unigram and bigrams whereas ROUGE-L helped capture the fluency. Reinforcement algorithm was used for minimizing the negative reward.

Lee et al. [70] used the DQN based Deep Learning models where the state referred to the incomplete sentences and the action refers to addition of sentences to the summary. Length limitation and the reward were used as the parameters. Kohita et al. [63] used the Q-Learning based RL model to create the summaries.



IV. DISCUSSION

Main tasks for which Machine Learning approaches have been used by various researchers for assisting in Text summarization observed from researches:

1. Redundancy

2. Sentence Extraction

3. Clustering

4. Sentiment Analysis

5. Optimization

6. Sentence Ranking

7. Ambiguity Removal

8. Noise Removal

9. Keywords identification

Evaluation Measures and Results:

The main focus of Machine Learning approaches is to find the important sentences by learning the weights of the features given in the dataset from the training dataset. The main evaluation measures which are used to find the efficiency and effectiveness of a particular Machine Learning approach are:

- Confusion Matrix: Assuming there are N Classes for prediction by the Machine Learning Model, it is a N * N table where the classification results are summarised. In the confusion matrix evaluations, mainly following terminologies are used:
- True Positive : It means the Predicted Class True

for the dataset is correct. It is the instance of correct classification by the algorithm

- True Negative: It means the Algorithm predicted the class to be True but the actual Class was False
- False Positive: It means the Algorithm predicted the class to be False but it's actual class was True
- False Negative: It means the Algorithm predicted the class to be False and the actual class was also False.
- The main objective of Machine Learning Algorithms is to maximize the True Positive Rate and minimize the False Positive Rate.

Precision: It is the ratio between True Positive and sum of True Positive and False Positive. It mainly calculates the amount of correctly predicted positives.

Recall: It is the ratio between True Positive and sum of True Positive and False Negative. It calculates the actual positives which were classified correctly.

F-Score: It is the harmonic mean between Precision and Recall. Higher is the F-Score, better is the model.

Accuracy: When the machine learning algorithms are applied to the dataset, it classifies some data points correctly while some wrong. Accuracy indicates the percentage of classifications that were done correctly by the approach.

ROC Curve: It is also known as Receiver Operating Characteristic Curve. It is the plot between True Positive Rate vs False Positive Rate.

True Positive Rate is the ratio between True Positive to the sum of True Positive and False Negative. True Positive Rate is also known as Sensitivity,

False Positive Rate is the ratio between False Positive to the sum of False Positive and True Negative.

AUC : It is the area under the ROC Curve. Larger is the area under the ROC Curve, better is the model. It is also known as specificity.

Entropy Loss: It is mainly used for logistic Regression models and other Neural Based Networks. Smaller is the function, better is the model.

Human Based Metrics:

- Informativeness
- Fluency
- Readability
- Conciseness
- Relevance
- Non Redundancy
- Sentimental Accuracy

Technique	Evaluation Results
SVM	0.3813 (ROUGE-1) (DUC)
Maximum Entropy	0.3748 (ROUGE-1) (DUC)
Naive Bayes	0.3762 (ROUGE-1) (DUC)
Deep Learning Approaches	0.15-0.45 (ROUGE)(CNN/DailyMail)
Reinforcement Learning	0.16-0.41 (ROUGE) (CNN/DailyMail)

Table 1: Evaluation measures

1. Pros and Cons:

This section lists down a few of the strengths and the limitations of Machine Learning Based approaches for the text summarization task. This section also discusses this in detail technique wise for SVM, Naive Bayesian, Logistic Regression, Random Forests, Decision Trees, K-NN, Linear Discriminant Analysis, Neural Networks and Deep Learning Models. It will help the researchers choose the technique according to the requirement and the availability of data set.

Strengths:

- Learn Non Linear Relations
- Can Identify outliers

- Identify the patterns
- Helps achieve automation
- Can handle multi-dimensional data

Limitations:

- Need for Large Dataset
- Need for Computational Resources
- Overfitting
- Result Interpretation
- Dependency on the quality of dataset

Technique	Pros	Cons
SVM	Performs good for high dimensional data Works good with the outliers also Works good for binary classification related problems	Take a long processing time when dataset is large Do not distinguish properly when classes are not clearly separable Selection of kernels and hyperparameters impacts the performance of the approach
Naive Bayesian	Scalable Fast Works good with multi-class problems Works well with multi-dimensional data	The algorithm works on the hypothesis of independence of variables but many times it is not applicable. Requires well populated training data as it is based upon the concept of conditional probability
Logistic Regression	Hyperparameters tuning is not required Simple Effective Scaling of features is not required	Does not work well with non-linear data Does not work well for correlated data
Random Forest	Works well for correlated data It produces less variance and error Works well with imbalance data	Difficult to understand

	Works well with data with missing values Handles overfitting problem very nicely and is less impacted with the outliers	
Decision Trees	Normalization of data is not required Handles the missing values nicely	Prone to overfitting Takes large time to come to the decision
K-NN	Simple to understand and implement Does not make any assumption about the data Evolves as the new data appears	Suffers with curse of dimensionality Does not work well with imbalance and is not very scalable Does not handle the outliers and
		missing values very nicely
Linear Discriminant Analysis	Good for multivariate distribution Compute confidence interval	Not good for multi collinear data
Neural Networks	Flexible and works good for both the regression and classification problem Works well for non linear data Fast prediction and are scalable	They are like black boxes where it is difficult to understand how the independent variables are impacting dependent variables Very expensive to apply on normal CPUs and need GPUs Relies very much on training dataset and thus are prone to overfitting and generalization

Deep Learning Models	They can generated additional features from the existing training dataset without human intervention Very good for those tasks which require lot of feature engineering Works well for unstructured data Can perform many complex problems simultaneously Scalable	Requires massive training dataset Demands a lot of computational power

Famous Datasets:

1. CNN/DailyMail -corpus: The dataset consists of newspaper articles. The dataset consists of 286,817 articles for the validation purpose whereas contains 11,487 articles for the testing purpose. The dataset contains on average 766 words and around 30 sentences per document. This dataset was created by Lin et al. []. This dataset is very good and a famous dataset due to the fact that it contains very high quality summaries in terms of grammar.
2. DUC Datasets: DUC is a conference which started from the year 2001 to encourage and recognize the efforts made for the summarization task. DUC releases the dataset consisting of english documents for the evaluation of the approaches created by the researchers. The same conference became TAC after 2008. Most of the documents available in these datasets are from newspaper headlines. DUC 2001 and 2002 dataset consisted of more than 200 documents and 50 to 200 word abstractive summaries while 2003 dataset consisted of 100 words abstractive summaries. TAC 2010 and TAC 2011 consisted of summaries with fixed 100 words length.
3. Gigaword: The corpus contains around 3.8 M training datasets. It also contains the headlines for the news.
4. WikiSum is the dataset with around 1000 documents from wikipedia and is widely being used for the evaluation of Multi Document summarization approaches.
5. ACL Anthology References
6. New York Times Dataset: It is a dataset having the documents from 1995 to 2008. It has been mainly used for evaluating the extractive text summaries.
7. NewsRoom
8. XSum

Challenges:

1. Seq2Seq Models which are widely used for the purpose of abstractive summarization suffers from the repetition and the semantic irrelevant sentences. Also most of the summaries which are generated lack in terms of grammar, and concise reflection of main ideas. More work on incorporating the tree structure to the existing LSTM and RNN models can be done. Hou et al. [].
2. Most of the Deep Learning Models utilise the information from the original text only but more efforts on incorporating the information from the external sources can improve the summaries generated. This external knowledge can be taken by using Lexical chains, or utilizing the Rhetorical Tree Structure, or knowledge graphs.

3. At present the Machine Learning Based approaches are based upon the extraction of sentences by considering the Feature Scores. But in future, the clause-level extraction can also be tried for getting better summaries in terms of coherence and conciseness.
4. Most of the research generates the generic summaries but more attention is required on generating the summaries which considers the user preference also.
5. Traditional approaches used for neural networks take longer training time, face the issue of overfitting of data, and number of hidden layers limit the learning capability.

Future Directions:

Lot of research has been done in this field. Different approaches have been developed by various researchers considering various parameters for the text. Many advanced algorithms like graph-based, Deep Learning models, Reinforcement Learning, Information Retrieval based approaches have been used to get better results in terms of not just quantitative measures but also qualitative measures. Thus. Now the research in this field is shifting from just extracting the important sentences and arranging them and thus this has raised a lot of challenges.

For most of the approaches where the artificial intelligence related approaches have been used, the features used are the traditional features which have been used by researchers a decade ago. Need to look at the features from word and sentence level is the need of time as it has been observed that better the features we choose and better we train them , better are the results that we obtain.

For getting the results closer to the natural form, it is important to have good and large datasets for the training purpose. Better is the training dataset, better the results are from artificial intelligence techniques. Even though we have very good datasets available, still there is a need for the training of deep learning approaches.

Most of the summaries are evaluated using the intrinsic measures use the ROUGE Scores or the BERT Scores where the focus is to measure the overlap of information between the system generated summaries and the human summaries but need is to have better measures which consider not only the overlap of information but also parameters like informativeness, fluency, quality in terms of coverage and cohesion. Few works have been done which considers the vocabulary along with the ROUGE scores by incorporating information from sources like WordNet, but more work is required in this side also. ROUGE Scores are good for the extractive summaries evaluation but not suitable for evaluating the abstractive summaries due to the reason that ROUGE scores are based upon the word overlap.

More Datasets are required than TUC and DUC datasets for better research in this field. CNN/ DailyMails dataset are popular among the deep learning community but there is a need for other datasets from various domains suitable for both the extractive and abstractive summarization.

Deep Learning models have been used extensively nowadays for the summarization task. If multi modality is also considered, summaries will be more rich and can include not only information from text but also from videos and images. Limited Labelled data also poses another challenge to the training of deep learning models both at the encoder and decoder side.

More research on utilization of reinforcement learning so as to capture the environmental dynamics is also the need of time. Reinforcement learning techniques can capture the user defined metrics in a better way.

CONCLUSION

With the availability of fast and easy access to the internet, the amount of information in it has increased exponentially which has resulted in the need of automatic text summarization systems for quick retrieval of relevant information. Automatic text summarization is an interesting area of the research and is used in various domains like movies, software, legal text documents, google search engine, etc. The advancement in Machine Learning and Deep Learning Models has eased the task of automatic summarization. The objective of the paper is to systematically arrange the works in this field using the artificial intelligence techniques and especially ML and DL. This survey will serve as the beginning to the researchers who are willing to do their research in this field as it systematically mentions the works done by various researchers according to the techniques and also chronologically mentions the work for the understanding of the trends. The classification of work was done according to the approach and the techniques used, challenges and the future research areas are also discussed in the paper. The datasets, techniques used for the research are also mentioned to have the better familiarity to the publicly available dataset for the interested researchers.

REFERENCES

[1] Hans Peter Luhn. ”The automatic creation of literature abstracts.” IBM Journal of research and development 2, no. 2 (1958): 159–165.

[2] H. P Edmundson. ”New methods in automatic extracting.” *Journal of the ACM* 16, no. 2 (1969):264–285.

[3] Ani Nenkova, and Kathleen McKeown. “A Survey of Text Summarization Techniques”. sMining Text Data (2012): 43-76.

[4] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Ranking Sentences for Extractive Summarization with Reinforcement Learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1747–1759, New Orleans, Louisiana. Association for Computational Linguistics.

[5] Mahak Gambhir and Vishal Gupta. 2017. Recent automatic text summarization techniques: a survey. Artif. Intell. Rev. 47 (1):1–66.

[6] Som Gupta and S.K Gupta. 2018. Deep Learning in Text Summarization. In International Journal of Computer Science and Information Security Vol 16.no 11. Pp 150-155

[7] C. Y. Lin. 1999. “Training a selection function for extraction,” In Proceedings of the eighth International Conference on Information and Knowledge Management, ACM: 55–62.

[8] C. Aone, M. Okurowski, and J. Gorlinsky. 1998. Trainable Scalable Summarization Using Robust NLP and Machine Learning. Proc. of the 17th COLING and 36th ACL: 62–66.

[9] Chuang Wesley T. and Yang, Jihoon. 2000. Extracting sentence segments for text summarization. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval: 152–159.

[10] John M. Conroy and Dianne P. O'leary. 2001. Text summarization via hidden Markov models. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01). Association for Computing Machinery, New York, NY, USA, 406–407.

[11] Smaranda Muresan and Evelyne Tzoukermann and Judith L. Klavans. 2001. Combining linguistic and machine learning techniques for email summarization. In Proceedings of the 2001 workshop on Computational Natural Language Learning - Volume 7 (ConLL '01). Association for Computational Linguistics, USA 19: 1–8.

[12] Amini, M.-R., Gallinari, P.: The Use of unlabeled data to improve supervised learning for text summarization. In: Proceedings of the 25*th* ACM SIGIR, pp. 105–112 (2002)

[13] Amini, M.R., Usunier, N., Gallinari, P. (2005). Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms. In: Losada, D.E., Fernández-Luna, J.M. (eds) Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg.

[14] M. Osborne.2002. Using maximum entropy for sentence extraction. In Proceedings of the ACL Workshop on Automatic Summarization: 1–8.

[15] Tsutomu Hirao and Hideki Isozaki and Eisaku Maeda and Yuji Matsumoto. 2002. Extracting important sentences with support vector machines. In Proceedings of the 19th international conference on Computational linguistics - Volume 1 (COLING '02). Association for Computational Linguistics, USA: 1–7.

[16] Neto J.L.and Freitas A.A. and Kaestner C.A.A. 2002. Automatic Text Summarization Using a Machine Learning Approach. Bittencourt G.,

Ramalho G.L. (eds) Advances in Artificial Intelligence. SBIA 2002. Lecture Notes in Computer Science. 2507: 205-215.

[17] Fattah, M.A. 2014. A hybrid machine learning model for multi-document summarization. Appl Intell 40: 592–600.

[18] Danushka Bollegala, Naoaki Okazaki, and Mitsuru Ishizuka. 2006. A Bottom-Up Approach to Sentence Ordering for Multi-Document Summarization. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 385–392, Sydney, Australia. Association for Computational Linguistics.

[19] Leite, Daniel Saraiva, Lucia Helena Machado Rino, Thiago Alexandre Salgueiro Pardo and Maria das Graças Volpe Nunes. "Extractive Automatic Summarization: Does more Linguistic Knowledge Make a Difference?" (2007).

[20] Kianmehr Keivan and Gao Shang and Attari Jawad and Rahman M. Mushfiqur and Akomeah Kofi and Alhajj Reda and Rokne Jon and Barker, Ken 2009. Text summarization techniques. Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services: 487–491.

[21] Mahmood Yousefi-Azar, Len Hamey, Text summarization using unsupervised deep learning, Expert Systems with Applications,Volume 68,2017,Pages 93-105, ISSN 0957-4174.

[22] Raut, P.S., Chaudhari, S.D., Waghole, V.B., Jadhav, P.U., & Saste, A.B. (2022). Automatic Evaluation of Descriptive Answers Using NLP and Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*.

[23] Silva Gabriel and Ferreira Rafael and Lins Rafael Dueire and Cabral Luciano and Oliveira, Hilário and Simske Steven J. and Riss, Marcelo. 2015. Automatic Text Document Summarization Based on Machine Learning. Proceedings of the 2015 ACM Symposium on Document Engineering : 191–194.

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. J. Artif. Int. Res., 16(1):321–357, June 2002.

[25] Lanyo, Kevin and Wausi, Agnes. 2018. A Comparative Study of Supervised and Unsupervised Classifiers Utilizing Extractive Text Summarization Techniques to Support Automated Customer Query Question-Answering. IEEE 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI): 88–92.

[26] M. Day and C. Chen, " Artificial Intelligence for Automatic Text Summarization," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, 2018, pp. 478-484, doi: 10.1109/IRI.2018.00076.

[27] Zawbaa H.M., El-Bendary N., Hassanien A.E., Kim T. (2011) Machine Learning-Based Soccer Video Summarization System. In: Kim T. et al. (eds) Multimedia, Computer Graphics and Broadcasting. MulGraB 2011. Communications in Computer and Information Science, vol 263.

[28] K. Kumar and D. D. Shrimankar, "F-DES: Fast and Deep Event Summarization," in *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 323-334, Feb. 2018, doi: 10.1109/TMM.2017.2741423.

[29] Ahmed, M. Data summarization: a survey. *Knowl Inf Syst* 58, 249–273 (2019). https://doi.org/10.1007/s10115-018-1183-0

[30] D. Taeho Jo, "Validation of Graph Based K Nearest Neighbor for Summarizing News Articles," *2019 International Conference on Green and Human Information Technology (ICGHIT)*, Kuala Lumpur, Malaysia, 2019, pp. 66-69, doi: 10.1109/ICGHIT.2019.00022.

[31] Gupta, Vivek, et al. "SumPubMed: Summarization dataset of PubMed scientific articles." *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*. 2021.

[32] Adem A. Karmis, Alper Derya, and Ali C. Begen. 2020. Metadata-based user interface design for enhanced content access and viewing.

Proceedings of the 11th ACM Multimedia Systems Conference (MMSys '20). Association for Computing Machinery, New York, NY, USA,379–381. Kupiec J. and Pedersen, J. O and Chen F. 1995. A trainable document summarizer. In Proceedings of the 18th ACM-SIGIR Conference, Association of Computing Machinery : 68-73.

[33] Pattanaik, A., Sagnika, S., Das, M., Mishra, B.S.P. (2019). Extractive Summary: An Optimization Approach Using Bat Algorithm. In: Hu, YC., Tiwari, S., Mishra, K., Trivedi, M. (eds) Ambient Communications and Computer Systems. Advances in Intelligent Systems and Computing, vol 904. Springer, Singapore.

[34] Larocca Neto and J.; Santos and A. D.; Kaestner and C.A.; Freitas and A.A. 2000 Document clustering and text summarization. In Proceedings of 4th Int. Conf. Practical Applications of Knowledge Discovery and Data Mining (PADD-2000) London: The Practical Application Company : 41-55.

[35] Mani I. and Bloedorn E. Machine Learning of Generic and User-Focused Summarization. In Proceedings of the Fifteenth National Conference on AI (AAAI-98). 1998):821-826.

[36] Yousefi-Azar M. and Hamey, L. 2002. Text summarization using unsupervised deep learning. Expert Systems with Applications.2002. 68: 93–105.

[37] Kaikhah, K. 2004. Text summarization using neural networks. Faculty Publications, Texas State University

[38] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. 2005. Learning to rank using gradient descent. Proceedings of the 22nd international conference on Machine learning (ICML '05). Association for Computing Machinery, New York, NY, USA: 89–96.

[39] Shardan, R. , and Kulkarni, U. 2010. Implementation and evaluation of evolutionary connectionist approaches to automated text summarization. Journal of Computer Science, 6 (11):1366–1376 .

[40] Hasselqvist, J. and Helmertz, N. and Kågebäck, M. 2017. Query-based abstractive summarization using neural networks. In: CoRR (2017).

[41] Ramesh Nallapati and Bowen Zhou and Cícero Nogueira dos Santos and Çağlar Gülçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to sequence rnns and beyond. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning: 280–290.

[42] Song, S. and Huang, H. and Ruan, T. 2019. Abstractive text summarization using LSTM-CNN based deep learning. Multimed Tools Appl 78: 857–875.

[43] Aakash Sinha, Abhishek Yadav, and Akshay Gahlot. Extractive Text Summarization using Neural Networks.

[44] . Zhang, J. E. Meng and M. Pratama. 2016. "Extractive document summarization based on convolutional neural networks," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society: 918-922.

[45] Junyang Lin and Xu Sun, Shuming Ma and Qi Su. 2018. Global Encoding for Abstractive Summarization. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers): 163–169.

[46] Deepa Anand and Rupali Wagh 2022. Effective deep learning approaches for summarization of legal texts. Journal of King Saud University - Computer and Information Sciences 34 (5): 2141-2150.

[47] Hasselqvist, J., Helmertz, N Kågebäck, M. 2017 . Query-based abstractive summarization using neural networks. In: CoRR.

[48] Atif Khan and Muhammad Adnan Gul and Mahdi Zareei and R. R. Biswal and Asim Zeb and Muhammad Naeem and Yousaf Saeed and Naomie Salim. 2020. "Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm", Computational Intelligence and Neuroscience. 14:1-14.

[49] T. Ayodele and R. Khusainov and D. Ndzi. 2017. "Email classification and summarization: A machine learning approach". IET Conference on

Wireless, Mobile and Sensor Networks (CCW SN07): 805-808.

[50] S. Mandal and P. Achary and S. Phalke and K. Poorvaja and M. Kulkarni. 2021. "Extractive Text Summarization Using Supervised Learning and Natural Language Processing,". International Conference on Intelligent Technologies (CONIT): 1-7.

[51] S. P. Singh, and A. Kumar and A. Mangal and S. Singhal.2016. "Bilingual automatic text summarization using unsupervised deep learning,". International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT): 1195-1200.

[52] S. Agarwal and N. K. Singh and P. Meel. 2018. "Single-Document Summarization Using Sentence Embeddings and K-Means Clustering,". International Conference on Advances in Computing, Communication Control and Networking (ICACCCT): 162-165.

[53] Rahul, ; Adhikar, Surabhi; Monika 2020. NLP based Machine Learning Approaches for Text Summarization. [IEEE 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) - Erode, India (2020.3.11-2020.3.13)] 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC): 535–538. https://towardsdatascience.com/pros-and-cons-of-various-classification-ml-algorithms-3b5bfb3c87d6 : Nov: 2022

[54] M. Okumura, Y. Haraguchi, and H. Mochizuki. 1999. Some Observations on Automatic Text Summarization Based on Decision Tree Learning (in Japanese). Proc. of the 59th National Convention of IPSJ (5N-2): 393– 394.

[55] I. Mani and E. Bloedorn. 1998. Machine Learning of Generic and User-Focused Summarization. Proc. of the 15th AAAI: 821–826.

[56] Chin-Yew Lin. 1999. Training a Selection Function for Extraction. Proc. of the 18th ACMCIKM: 55–62.

[57] Congbo Ma and Wei Emma Zhang and Mingyu Guo and Hu Wang, and QUAN Z. Sheng. 2022. Multi-document Summarization via Deep Learning Techniques: A Survey. ACM Comput. Surv.

[58] Tian Shi, Yaser Keneshloo, Naren Ramakrishnan, and Chandan K. Reddy. 2021. Neural Abstractive Text Summarization with Sequence-to-Sequence Models. ACM/IMS Trans. Data Sci. 2(1): 1- 37.

[59] Sheng-Luan Hou and Xi-Kun Huang and Chao-Qun Fei and Shu-Han Zhang and Yang-Yang Li and Qi-Lin Sun and Chuan-Qing Wang 2021. A Survey of Text Summarization Approaches Based on Deep Learning . Journal of Computer Science and Technology:.633-663

[60] K. Wong and M. Wu, and W. Li. 2008 . Extractive summarization using supervised and semi-supervised learning. In Proceedings of the 22nd International Conference on Computational Linguistics: 985–992.

[61] K. Kandasamy and P. Koroth. 2014. “An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques,” IEEE Students’ Conf. Electr. Electron. Comput. Sci. SCEECS : 1–5. https://paperswithcode.com/dataset/cnn-daily-mail-1. Sep 2022.

[62] K. Knight and D. Marcu. 2000. “Statistics-based summarization- step one: Sentence compression”. In Proceedings of the Seventeenth National Conference of the American Association for Artificial Intelligence: 703–710.

[63] https://cs.nyu.edu/~kcho/DMQA/. Sep 2022

[64] https://duc.nist.gov/data.html. Sep 2022

[65] https://deepai.org/dataset/gigaword. Sep 2022

[66] https://www.kaggle.com/datasets/cvltmao/wikisum-ranked. Sep 2022

[67] Alquliti, Wajdi & Binti, Norjihan. 2019. Convolutional Neural Network based for Automatic Text Summarization. International Journal of Advanced Computer Science and Applications. 10(4): 200-211.

[68] Suleiman D, Awajan A.2020. Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges. Math Probl Eng: 1-29.

[69] K. Lopyrev, Generating news headlines with recurrent neural networks. 2015, https://arxiv.org/abs/1512.01712: 1-9

[70] Alexander M. Rush, Sumit Chopra, Jason Weston. A Neural Attention Model for Abstractive Sentence Summarization. Proceedings of EMNLP 2015. arXiv:1509.0068

[71] Sumit Chopra, Michael Auli, and Alexander M. Rush. 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: 93–98.

[72] James Bradbury, Stephen Merity, Caiming Xiong, Richard Socher. "Quasi-Recurrent Neural Networks". ICLR 2017. arXiv:1611.01576.

[73] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gủlçehre, and Bing Xiang. 2016. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning: 280–290.

[74] Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou. 2017. Selective Encoding for Abstractive Sentence Summarization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics 1:1095–1104.

[75] Cao, Ziqiang and Wei, Furu and Li, Wenjie and Li, Sujian. 2017. Faithful to the Original: Fact Aware Neural Abstractive Summarization. Proceedings of the AAAI Conference on Artificial Intelligence: 1-8.

[76] Tian Cai and Mengjun Shen and Huailiang Peng and Lei Jiang, and Qiong Dai. 2019. Improving Transformer with Sequential Context Representations for Abstractive Text Summarization. In Natural Language Processing and Chinese Computing: 8th CCF International Conference: 512–524.

[77] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics 1: 1073–1083

[78] Song, S. and Huang, H. and Ruan, T. 2019. Abstractive text summarization using LSTM-CNN based deep learning. Multimed Tools Appl 78: 857–875 .

[79] Al-Sabahi Kamal and Zuping Zhang and Kang, Yang. 2018. Bidirectional Attentional Encoder-Decoder Model and Bidirectional Beam Search for Abstractive Summarization. Computational Linguistics: 1-9.

[80] Wojciech Kryściński, Romain Paulus, Caiming Xiong, and Richard Socher. 2018. Improving Abstraction in Text Summarization. In Proceedings of the 2018 conference on Empirical Methods in Natural Language Processing: 1808–1817.

[81] Yao K and Zhang L and Du D and Luo T and Tao L and Wu Y. 2020. Dual Encoding for Abstractive Text Summarization. IEEE Trans Cybern 50(3): 985-996.

[82] Watanabe, Taro and Sumita, Eiichiro. 2002. Bidirectional Decoding for Statistical Machine Translation. Bidirectional decoding for statistical machine translation:1-7.

[83] Yang Liu and Mirella Lapata. 2019. Text Summarization with Pre Trained Encoders. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): 3730–3740.

[84] Wang Q, Liu P, Zhu Z, Yin H, Zhang Q, Zhang L. A Text Abstraction Summary Model Based on BERT Word Embedding and Reinforcement Learning. Applied Sciences. 2019; 9(21):4701-4709.

[85] Elozino Egonmwan and Yllias Chali. 2019. Transformer-based Model for Single Documents Neural Summarization. In Proceedings of the 3rd Workshop on Neural Generation and Translation: 70–79.

[86] Z. Li, Z. Peng, S. Tang, C. Zhang and H. Ma, 2000. "Text Summarization Method Based on Double Attention Pointer Network," in IEEE Access 8: 11279-11288.

[87] Gyoung Ho Lee and Kong Joo Lee. 2017. Automatic Text Summarization Using Reinforcement Learning with Embedding Features. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 193–197, Taipei, Taiwan. Asian Federation of Natural Language Processing.

[88] Seonggi Ryang and Takeshi Abekawa. 2012. Framework of Automatic Text Summarization Using Reinforcement Learning. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 256–265, Jeju Island, Korea. Association for Computational Linguistics.

[89] Ryosuke Kohita, Akifumi Wachi, Yang Zhao, and Ryuki Tachibana. 2020. Q-learning with Language Model for Edit-based Unsupervised Summarization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 470–484, Online. Association for Computational Linguistics.

[90] Yulong Pei, Wenpeng Yin, Qifeng Fan, and Lian'en Huang. 2012. A Supervised Aggregation Framework for Multi-Document Summarization. In *Proceedings of COLING 2012*, pages 2225–2242, Mumbai, India. The COLING 2012 Organizing Committee.