A Combined Review of Text Summarization and Topic Modeling

Monika Singh

Department of Computer Science, Indira Gandhi Delhi Technical University for Women, New Delhi, India

imonikasingh1999@gmail.com

Abstract

Nobody can picture their daily lives today without a smartphone or the Internet. It is now a crucial home. How can I choose the finest option out of several options at the same price and identify which ones are genuine? Before making an online purchase, every consumer examines reviews. But not everyone enjoys reading these lengthy reviews. Therefore, there must be a way to distill a lengthy description into a short, multi-word phrase that expresses the same notion. To do this, there are numerous NLP strategies available such as topic modeling and text summarization. The research in the area of text summarization and topic modeling that was published between the years 2002 and 2022 is reviewed in-depth and methodically in this paper. The analysis's findings give a detailed explanation of the issues and trends that are the subject of their text summarization and topic modeling research; describe the approaches and strategies that are frequently used by researchers for method development and comparison.

Keywords

Text Summarization, Topic Modeling, Customer Review

1. INTRODUCTION

As the modern world is evolving day by day along with the technologies, with that comes the dependency on our machines.Everyone wants their work to be done as fast as possible and with minimum chance of failure. With the latest world , we came across online shopping, moving along with trends regardless of whether we acknowledge it or not, plays a significant part in our lives. And as we can observe the Covid time when we were at home and not being able to go anywhere, at that time we wanted all our shopping to be done without going out, so it kinda gave exposure to online shopping a bit more. With that comes a challenging part which product should i buy as just for a simple need there are hundreds of products available in hundreds of brands. So, selecting a good product with excellent reviews is important, we can observe there are a lot of reviews posted for a product on the websites and going through every product is kind of a hectic task to do which can result in changing our mind for shopping. So here comes the role of our technologies which makes this task convenient for us by providing various approaches such as Natural Language Processing which comes with the method of topic modeling and text summarization which helps in making the content of review short and precise to understand.

Topic modeling is basically the practice of automatically identifying subjects in textual material and discovering underlying patterns represented by textual information. There are many methods to implement topic modeling for organizing the data. The latent Dirichlet distribution (LDA), a descriptive statistical method that divides a set of observations into unsupervised groups and explains why particular feature data are associated in each group, is one of the most efficient ways to execute topic modeling. The text of a paper on a certain topic can also benefit from classification.

Text summarization is another crucial component of NLP. We may condense our writing to a few lines without losing its significance by removing extraneous information and formatting the remaining content into a more concise, semantic text structure. The two primary techniques for text summarization in NLP are extractive and abstract summarization.

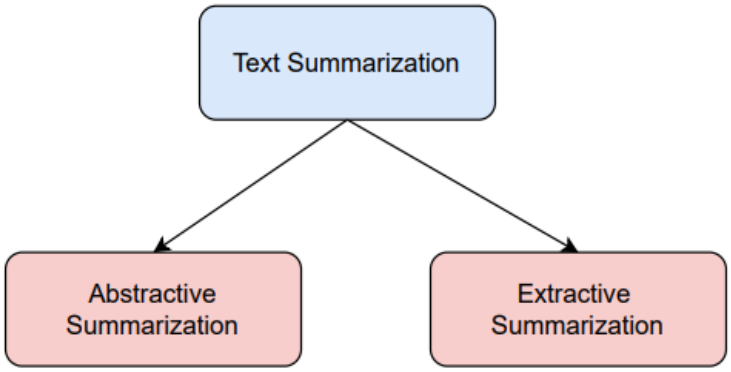

Fig. 1 Text Summarization Methods

1.1 Extractive Summarization

Extraction summarization makes an effort to condense articles by selecting substantial terms or expressions from the original content and merging different pieces of material to produce a compressed version. Afterward, a summary is made using these sentences that were extracted.

1.2 Abstractive Summarization

Contrary to extraction, this method depends on the capacity to parse and condense portions of a document utilizing sophisticated natural language techniques. Considering that abstract machine learning algorithms may produce fresh words and phrases to accurately reflect the sense of the original text. If used appropriately, these abstractions can aid in overcoming grammatical faults in deep learning exercises.

In our paper, we'll cover the research on topic modeling and text summarizing that has been done. A discussion of the literature review, which provides a summary of the studies being conducted in these disciplines, will be covered in section 2. The numerous text summarizing and topic modeling models that have been applied in various research throughout the years will be covered in section 3. The limitations that have been noted in earlier studies will be covered in section 4. The information about the datasets that are frequently utilized in these sectors will be provided in section 5. An overview of the evaluation matrices, which are used to assess the performance of different models, will be given in section 6. In section 7, we will talk about the difficulties encountered in earlier studies and outline potential areas for further research.

1.3 Datasets Used

1.3.1 DUC-2004 DATASET FOR DOCUMENT UNDERSTANDING CONFERENCES

This Document Understanding Conferences DUC dataset has 500 news articles [4] [17], and their summaries take up 75 bytes. The DUC dataset's main advantage over competing datasets is that the summaries are created by human authors and that there are likely one or more summaries per article. The Document Understanding Dataset DUC [19] is used in conjunction with other existing datasets [11]–[12] because it cannot be utilized to train models with a high number of parameters.

1.3.2 NEW YORK TIMES DATASET

Suggested and released major article sets between 1995 and 2008 are included in New York Times NYT dataset [2]. For extractive systems, this dataset is helpful in determining how abstract their system is using the NYT [30].

1.3.3 DATASET FROM CNN AND DAILY MAIL

The dataset used to evaluate several text summarizations [1]–[3] is the CNN/DailyMail dataset [6] [7] by Nallapati et al [11] [25]. This dataset includes 780 tokens worth of internet news items as well as a summary with 57 tokens worth of phrases that is also paired. The modified version of the CNN/DailyMail dataset is an additional version, has around 287,230 training pairings, 13,370 validation pairs, and 11,492 test pairs.

1.3.4 GIGAWORD DATASET

The Gigaword [13] [14] news article collections are helpful for creating text summaries. The original article does not contain summaries that are attached to it. However, several earlier works used this dataset subset and created two sets of summaries based on each headline's first line as well as the article [19].

Datasets	Descriptions
CNN and Daily Mail dataset	● utilized to assess the creation of text summaries ● Contains summaries of multiple-sentence web news stories.
New York Times dataset	● extensive selection of articles ● The system of abstraction that is most frequently used for extractive systems is evaluated first.
Document Understanding Conferences DUC-2004 dataset	● includes news story summaries. ● The fact that the article contains summaries in multiples of one is the main benefit. ● able to be combined with other datasets
Amazon review dataset	● Product reviews and information are included. ● Contains product reviews, links, and metadata
Gigaword dataset	● collection of news items used for summarizing ● The main article is free of summaries that are attached to it.
Newsroom dataset	● a large-scale dataset for the creation of text summaries ● consists of a combination strategy that is extractive and abstract.

Table 1 Used Datasets Descriptions

2. LITERATURE REVIEW

Ref No.	Authors	Title	Year	Description	Methods
[1]	Vidyasagar Potdar and Yee W. Lo	A review of opinion mining and sentiment classification framework in social networks	2009	analyses the many procedures used and the existing research on sentiment evaluation and data mining of online customer comments and reviews.	Sentiment Classification, opinion mining
[2]	Bing Liu and Minqing Hu	Mining and summarizing customer reviews	2004	done in the following steps: (1) locating product attributes that customers have mentioned; (2) determining if each reviewer's opinion is positive or negative; and (3) summarizing the results.	Feature-Based Summarization
[3]	KeXu, Hui Zhang, Junjie Wu, Yuan Zuo, and Junjie Wu	Complementary Aspect-Based Opinion Mining	2018	Across asymmetric collections, a topic model for complementary aspect-based opinion mining	Max Ent-LDA model
[4]	RN Ganesh Kumar and JN Madhuri	Extractive Text Summarization Using Sentence Ranking	2019	It is shown how to use statistics to extract text from a single document and summarize it.	Text summarization techniques that are both abstract and extractive
[5]	G. Dorffner, M. Samwald, and M. Moradi	Deep contextualized embeddings for quantifying the informative content in biomedical text summarization	2019	Research is being done to demonstrate how contextualized embeddings are created by employing a sophisticated bidirectional language model to determine how informative a sentence is while summarizing biological information.	BERT model
[11]	Xiang, B., Gulcehre, C., and Nallapati, R.	Abstractive text summarization using sequence-to-sequenc e rnns and beyond	2016	models that handle important summarizing issues that the basic architecture does not fully simulate, such as	RNN encoder with GRU decoder

				modeling crucial words, creating words that are uncommon or unseen during training and memorizing the order of sentence-to-word structure	
[6]	L. Canini, S. Benini, N. Adami, A. Signoroni, and R. Leonardi	A free Web API for single and multi-document summarization	2017	Different text analysis methods, such as topic modeling, sentence clustering, and keyword and entity extraction, produce competitive SoA outcomes.	ROUGE measure
[7]	Tateishi, K., Ya Yamanishi, S. Morinaga, and Fukushima	Mining Product Reputations on the Web	2002	determining whether a remark is an opinion or not and whether it is a good or negative view, create in advance syntactic and linguistic guidelines.	Text mining can be divided into four different categories. 1) defining terms, 2) co-occurring words, 3) usual expressions, and 4) Comparison of connection between various goal categories
[8]	Wilson, T., Wilson, Wiebe, and Litman	Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering	2003	create a mechanism for automatically creating opinion-based summary depictions and how they could be applied to a range of multi-perspective question answering challenges.	information-extraction approach
[9]	H. A. Caldera and I. K. C. U. Perera	Aspect based opinion mining on restaurant reviews	2017	These reviews can be analyzed using sentiment analysis, which categorizes issues that are either uplifting or depressing.	Aspect-based, sentence-based, and document-based are the three different stages of opinion mining.
[10]	V. S. Rekha and P. V. Rajeev	Recommending products to customers using opinion mining of	2015	To compare and provide recommendations for online-sold products, develop a web-based	Opinion mining and naive - Bayes

		online product reviews and features		prototype system. The polarity of reviews was tested using Naive Bayes classification, and reviews were automatically read using natural language processing.	
[12]	Huang, S., Zhou, M., Zhao, T., Yang, N., Wei, F., Zhou, Q., and	Neural document summarization by jointly learning to score and select sentences	2018	a complete neural network architecture for deep document synthesis combining learning notation and sentence selection	LSTM with selective gate network
[13]	Z. Wang, L. Bing, W. Lam, and P. Li	Deep recurrent generative decoder for abstractive text summarization	2017	The quality of the target summaries is increased by using a repeated latent random pattern to learn the latent structural information included in the summaries.	RNN with an Attention mechanism
[14]	Chopra, S., Weston, Rush, A.M.	A neural attention model for abstractive sentence summarization	2015	uses input text to produce each word in the summary using a local attention-based model.	Complete RNN encoder decoder
[15]	Liu, P.J., See, A., and Manning, CD	Get to the point: Summarization with pointer generator networks	2017	The sequence - to - sequence attention model is improved in two different ways by the new design that is put forth.	LSTM with pointer generator network
[16]	Singh, G., Parikh, and Khatri, C.	Abstractive and extractive text summarization using document context vector and recurrent neural networks	2018	For a better generalization, the Seq2Seq model should be run using context data in the first input step.	Seq2Seq models using RNNs

Table 2 Literature Review

The article named "Exploring Opinions and Emotions classification framework in a social network" was suggested by Yee W. Lo and Vidyasagar Potdar. Al. recommended that consumers might share their comments regarding the products they have purchased or the services they have received from other businesses on the Internet .[1]

In contrast to the conventional text summary, the summary task suggested by Minqing Hu and Bing Liu exclusively concentrates on the quality of a product that users have remarked on, regardless of whether such remarks are favorable or unfavourable. [2]

By Yuan Zuo, Junjie Wu, Hui Zhang, Deqing Wang, and KeXu, "Complementary Aspect-based Opinion Mining." According to Yuan Zuo and others, aspect-based concept exploration is a process of discovery. A product itself could be the focus of remarks on a particular topic. The Max Ent-LDA model for cross-assembly with automated labeling (CAMEL) is used to assemble evaluations based on appearance. The primary CAMEL discoveries, the production procedure, and a rough CAMEL reverse mining model. [3]

Work on JN Madhuri and R Ganesh Kumar's "Extractive Text Summarization Using Sentence Ranking." When they used sentence classification with Python 3.6 and NLTK in their article, They discovered that the content comprised more than 8 sentences. Finally, audio is created from the embedded content. [4]

Deep contextualized embeddings for quantifying the informative information in biomedical text summarization is the title of an essay by M. Samwald, Moradi, and M. G. Dorffner uses datasets and the BERT model test to support past studies on physical development. When they came to, they realized that if 30% of the original text is shown, the size renders correctly. This method has been found to work with biomedical texts .[5]

M suggested a freeWeb API for summarizing both individual and many documents.A Signoroni, N. Adami, L. Canini, S. Benini, and R. Leonardi uses word processing, math, and translation. In the DUC2001 dataset, it uses Python libraries like scikit training and the nltk library, and it has demonstrated that it is more successful at ROUGE computations than at multi-text summaries.[6]

The ratings of several objects are contrasted by Morinaga et al. A category that allows you to learn how popular a particular product is Not at all; it doesn't replace reviews. Product features that reviewers have touched on in their reviews. Although there are several expressions that are frequently used in relation to reputation, examples include "doesn't work," "benchmark result," and "no problem.". [7]

Discussion on the gathering of data based on opinions by Cardi et al. The objective is to obtain a broad picture of the opinions. Answer the question. In a nutshell, they support using the "Scenario model" to represent opinions. an opinion-based document or group of documents. [8]

I. K. C. U. Perera and H. A. Caldera are the authors of the essay "Aspect Based Opinion Mining on Restaurant Reviews." I. K. C. U. Perera et al. advise most clients to use the internet for everything, including food and hospitality management, without much success. The foundations of leadership depend greatly on these emotions. [9]

In their paper "Recommending Products to Customers Using Opinion Mining of Online Product Reviews and Features," P. V. Rajeev and V. S. Rekha demonstrates how to use naïve Bayesian classification to identify the intensity of reviews and natural language processing to automatically scan reviews. [10]

3. TOPIC MODELING AND TEXT SUMMARIZATION TECHNIQUES

The text summarizing and topic modeling techniques that are currently in use will be thoroughly reviewed in this section.

3.1 Text Summarization Models

As discussed in Section I, now we are familiar with the types of text summarization. We observed from the published research work that some methods that are highly used in performing text summarization and we have provided the overview of the methods in this section.

3.1.1 BERT Model

RNNs and other neural networks' long-term dependencies are a drawback that BERT (Bidirectional Transformer) eliminates. This pre-trained model is naturally bidirectional. When BERT analyzes text, it uses a transformer, an attentional mechanism, to identify the contextual relationships between words (or subwords). In its default configuration, the converter has two distinct processes: a text-reading encoder

and a decoder that generates predictions about activity. Since the goal of BERT is to offer a linguistic model, all that is required is an encoder mechanism. [19] Using contextualized attachments produced by bidirectional encoder representations from model transducers, M. Moradi et al.'s proposed summary method summarizes information.[5] Figure 2 depicts the BERT model's flow. For many sentences, the BERT model has been upgraded to produce inline sentences. Add the [CLS] token before the start of the first sentence to accomplish this. A sentence vector is then generated for each sentence. The phrase vectors are then given multiple layers, which makes it simpler to capture data at the document level. The final sum prediction is compared to the true one in order to train the sum levels and the BERT model. [26]

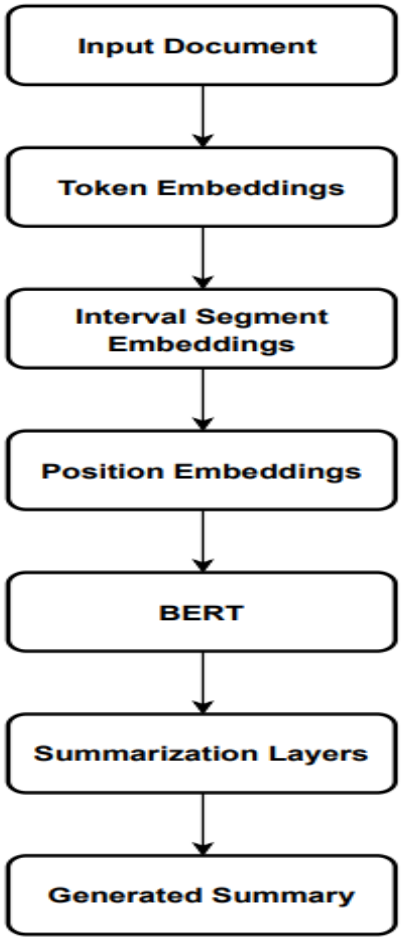

Fig 2 BERT Model Flow

Given that the model was the output vectors of a masked model that was trained—rather than the sentences—are tokenized. Instead of including a group of sentences like earlier extractive summaries, it leverages inclusions to distinguish between various sentences and only includes two labels, namely sentences A and B. The appropriate summaries are produced by modified versions of these embeddings. [11]

3.1.2 RNN

RNNs feature a built-in memory. When dealing with models that deal with data in a sequential way, these memory units are quite helpful. Other network topologies base a point's decision solely on the input and output that are currently available. In contrast, the previous output from an RNN is used as input for the

following calculation.[30] RNN utilizes the idea of back propagation. Gradient descent is used to lower the error after evaluating the error's partial mathematical derivation with respect to weight. RNN is applied using Back Propagation Through Time (BPTT). With unrolled RNN, it is used. Each time, it calculates the error and modifies the weights. Unrolled RNN is shown in the figure below.The input in this case is denoted by xt and ranges from 0 to time t. For each value of t, the outcome is ht. The letter A is used to symbolize neural networks.

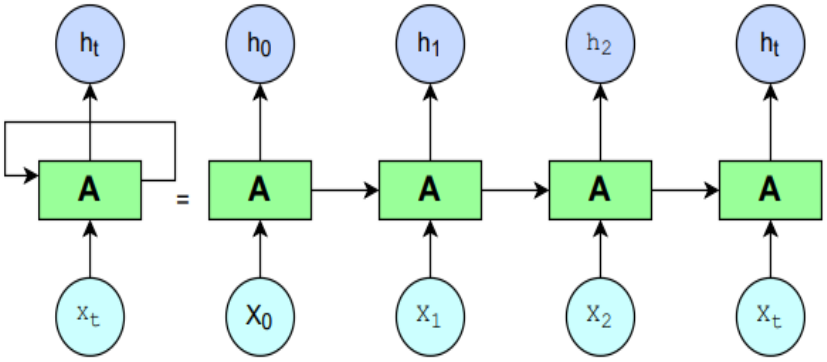

Fig 3 Unrolled RNN

However, RNN has two drawbacks. Gradient values will rise if the partial derivative values are greater than enormous, which results in bursting gradients. Similar to this, if the partial derivatives are extremely small, the gradient's values will decrease and eventually vanish. The inputs are stored for a longer time using LSTM networks.

3.1.3 LSTM

Short-term long-term memory refers to the components that a recurrent neural network's (RNN) layers are composed of (LSTM). A cell, an input state, an output state, and a forgotten state help compensate for an LSTM structure. A crucial component of the LSTM is the cell that aids in storing values for a predetermined amount of time. The capacity for memory gives rise to the term "long-term memory" (LSTM).[12] The three gates serve as current controllers that manage how values are transferred through the LSTM. These gates can activate a weighted sum by acting like neurons in a neural network. This model is distinguished by a longer short-term memory retention period.

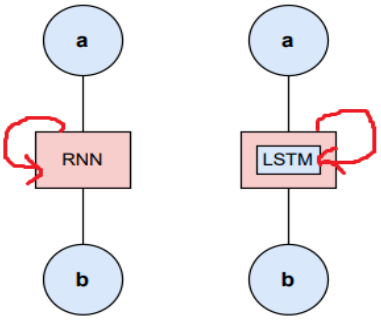

Fig 4 General representation of the RNN cell and the LSTM cell

The three fundamental inputs to the LSTM are the present input, the previous memory state, and the prior hidden state.The outputs of the LSTM are the current hidden and memory state. Each input vector, each latent vector, and their combined input sources of candidate gates, forgotten gates, input gates, and output

gates are used to create a set of output feature vectors from every of the authorized ports. Current status of the memory cell is part of the input, which the forget gate can modify. [15] The memory will forget its prior state if the forgetting gate's value is zero; else, pay heed. This is a straightforward method of describing long-term memory.

3.1.4 Sequence-To-Sequence (Seq2seq)

Understanding Seq2seq modeling is crucial to comprehending how various text summarizing techniques are implemented [28].

With the Seq2Seq method, a model is trained to convert a sequence from one domain to another. The seq2seq structure is used for things other than machine translation, like translating a German operator into English [29]. There are many uses for AI models that can answer questions. It's primarily employed to produce new texts. [16]

The target or output sequence in seq2seq must often be predicted using the whole input sequence. This specifically happens when the input and output sequence lengths are incompatible. Layers of encoders and decoders make up the majority of these systems. Encoder layer information is used to construct states, which the decoder layer then uses as a context or a condition. Using the target sequence's prior characters as a guide, the decoder layer forecasts the following characters.If the decoder is given targ(.....t), the input sequence's context, it will anticipate targ(t+1.....).[28] Separate input sequences, or unknown sequences, are decoded using a different technique.The encoder is the first place where input is sent before state vectors are obtained. The decoder is then fed a state vector with the target character of size 1, after which the newly detected character is sampled and attached to the target. The decoder gets its data from the encoder's state vectors in known sequences.

3.2 Topic Modeling Models

In earlier studies various types of techniques have been used for implementing topic modeling. We will provide an overview of some of the techniques.

3.2.1 Latent Semantic Analysis

LSA refers to methods or processes for NLP. Building a text that is displayed in vector form that produces in-depth semantics is the aim of latent semantic analysis (LSAprimary). In order to identify the most helpful linked terms, vector representation (LSA) calculates how similar the texts are to one another. LSA, which is now used for information extraction, was once known as Latent Semantic Indexing (LSI). [32] Due to this, only a small percentage of the numerous feature articles were chosen. To provide methods like keyword matching, weighted keyword matching, and vector rendering depending on a word's occurrence in a document, an LSA should comprise a variety of parts. Additionally, LSA uses SVD to arrange info.

All reductions to vector spaces are calculated and transformed by the SVD method using matrices. The reduced vector space is also calculated, and the elements are sorted from most to least significant. The most important premise in LSA is used to determine the text significance; otherwise, the least significant assumption is disregarded during the assumption. If two words have a similar vector, searching for words with a high rate of similarity will take place.[35] Collect a sizable amount of pertinent text, then divide it up into documents in order to describe the LSA's most important steps first. Second, create a a matrix of concepts and texts' co-occurrences. Include a document-specific n-dimensional vector, cell names like content x, words y, and m as the spatial values of terms. The next step involves whetting and computing each cell. The computation of all diminutions and creation of three matrices will ultimately rely heavily on SVD.

3.2.2 Probabilistic Latent Semantic Analysis

Probabilistic latent semantic analysis (PLSA), a Bag-of-Word (BOW) text analysis method, employs a probabilistic framework to find instances of meaningful phrases in corpora. Hoffman was the one who created it [39].

The Aspect model was the first statistical model to uncover the textual item matrix of the corpus, linguistic accompaniment. It is predicated on the idea that every word in a text originates using only one subject. Additionally, different phrases in a text might originate across a range of themes. Each document is depicted as a set of combinations of proportions for those components of a mixture in order to reduce it to a probability distribution across a specified range of topics.

3.2.3 Latent Dirichlet Allocation

The combined PLSA and LSA models, which hitherto took word and text interchange into account, were replaced by the Latent Dirichlet Allocation (LDA) model. Every set of interchangeable random variables has a representation as a mixed distribution, often an infinite mixture, according to the classical representation theorem, which was established in 1990. [40].

With its foundation in statistical (Bayesian) topic models, the Latent Dirichlet Allocation text mining algorithm is one of the most popular. An attempt to replicate the writing process is made through the generative model LDA. It then makes an effort to create a document on the specified subject. Other kinds of data may also be subject to it.Numerous LDA methods have been developed, such as supervised topic modeling techniques, text analysis, author and topic temporal analysis, LDA-based bioinformatics, and latent Dirichlet co-clustering [41], [42].

The fundamental notion behind the approach is that each text is described as a collection of topics, where each subject represents a discontinuous probabilistic model, each of which establishes the likelihood that a specific word will appear in that topic. A document can be summarized using these topic probabilities. A "document" in this context is simply other than the number of words and theme, the words are just thrown together in a "bag of words."

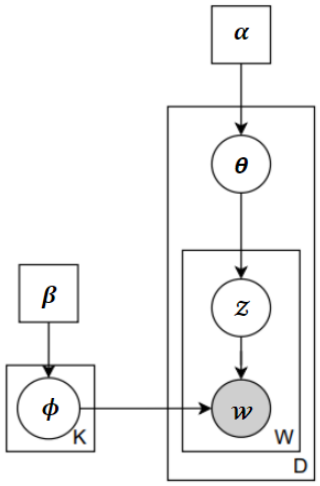

Fig 5 LDA is depicted in a graphical manner.

The K latent themes that define a multinomial distribution throughout the W word vocabulary are combined to model each of the D documents using LDA. In Figure 5, the LDA model is graphically shown. The main steps in the generation of an LDA are as :

With each time Nj appears in document j

i) Choose a topic zij ~ Mult(θj)
ii) Choose a word xij ~ Mult(φzij)

where topics and a topic φk has multinomial parameters with Priors to Dirichlet for subjects in document θj. [41]

4. LIMITATION

The methods utilized to accomplish text summarization and topic modeling have a few drawbacks that we have found. We will talk about the restrictions that are in place in this section.

4.1 Topic Modeling

- It is challenging to locate and count the subjects.
- Operating it and determining the loading values' probabilistic significance are difficult.
- At the document level, the PLSA does not support probabilistic modeling.
- Modeling relationships between subjects that can be resolved with the CTM method is lost.
- Numerous calculations are necessary, and the themes contain numerous broad phrases.

4.2 Text Summarization

- Manual summary creation requires human intervention.
- Both during the training phase and during the application phase, neural networks are slow.
- The net's decision-making process is difficult to comprehend.
- Understanding the net's decision-making process is challenging.
- It requires human interference to gather training data.
- domain-specific, so the training phase needs access to an external domain-specific corpus.
- The disadvantage is that language considerations are not made.

5. EVALUATION MATRICES

To evaluate the effectiveness of the construction model, there are numerous evaluation matrices available. This part will cover ROUGE for performance evaluation and Human Evaluation.

5.1 Human Evaluation

Automation of the analytical process is challenging since human opinion on what makes a "good" outline frequently varies substantially. Even though manual analysis is employed, Because of how labor-intensive and time-consuming it is, it necessitates reading both the source materials and the summaries. Other challenges include those with coherence and coverage.

5.2 ROUGE

A candidate abstract is compared to a group of guidelines abstracts using the n-gram ROUGE-N recall test. Follow these procedures to calculate ROUGE-N :

$$\text{ROUGE-N} = \frac{\Sigma s \varepsilon reference\_summaries\, \Sigma N-grams\, Countmatch(N-gram)}{\Sigma s \varepsilon reference\_summaries\, \Sigma N-grams\, Count(N-gram)}$$

where n denotes length n. Grams The counter match is the most N grammes possible that is compatible with the reference digests and candidate digests. The count is a way of expressing how many n-grams are present in the compound reference sentence (n-grams).

6. CHALLENGES AND FUTURE WORK

It is challenging to evaluate summaries (manually or automatically).The fundamental issue with scoring is that there is no set criteria by which to compare system scores.The system has the ability to produce a summary that is wholly distinct from any human-produced summary to approximate a right conclusion, making it impossible to define what exactly makes a correct summary. The issue with content selection is still unresolved[23]. Due to the wide range of individual personalities, subjective writers may have selected entirely different sentences. The process of combining two distinct sentences with distinctly dissimilar word meanings is known as paraphrasing. And sometime while doing the online shopping user can wish for checking the review summary topic wise. Keeping these challenges in mind , future study can be done on bringing topic modeling and text summarization together with better performance and accuracy in order to provide better convenience to the user while shopping.

7. CONCLUSION

The objective is to provide an in-depth analysis and comparison of various methodologies and strategies for the topic modeling and text summarizing processes.Additionally, it included a general review of the many strategies and datasets that can be utilized to swiftly and accurately summarize lengthy texts through the automatic generation of textual and thematic summaries. Despite substantial research on subject generalization and modeling studies, there is still much to be done. It has been less important over time to summarize and develop themes for scientific articles in favor of adverts, blogs, emails, and news. In many situations, simply eliminating phrases has led to satisfactory outcomes.

REFERENCES

[1] Lo, Valencia & Potdar, Vidyasagar. (2009). A review of opinion mining and sentiment classification framework in social networks. 396 - 401. 10.1109/DEST.2009.5276705.

[2] Hu, Minqing & Liu, Bing. (2004). Mining and summarizing customer reviews. KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 168-177. 10.1145/1014052.1014073.

[3] Y. Zuo, J. Wu, H. Zhang, D. Wang and K. Xu, "Complementary Aspect-Based Opinion Mining," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 2, pp. 249-262, 1 Feb. 2018, doi: 10.1109/TKDE.2017.2764084.

[4] J. N. Madhuri and R. Ganesh Kumar, “Extractive Text Summarization Using Sentence Ranking,” 2019 Int. Conf. Data Sci. Commun. IconDSC 2019, pp. 1–3, 2019, doi: 10.1109/IconDSC.2019.8817040.

[5] M. Moradi, G. Dorffner, and M. Samwald, “Deep contextualized embeddings for quantifying the informative content in biomedical text summarization,” Comput. Methods Programs Biomed., vol. 184, p. 105117, 2020, doi: 10.1016/j.cmpb.2019.105117.

[6] M. Mauro, L. Canini, S. Benini, N. Adami, A. Signoroni, and R. Leonardi, “A freeWeb API for single and multi-document summarization,” ACM Int. Conf. Proceeding Ser., vol. Part F1301, 2017, doi: 10.1145/3095713.3095738.

[7] Morinaga, S., Ya Yamanishi, K., Tateishi, K, and Fukushima, T. 2002. Mining Product Reputations on the Web. KDD’02

[8] Cardie, C., Wiebe, J., Wilson, T. and Litman, D. 2003. Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. 2003 AAAI Spring Symposium on New Directions in Question Answering.

[9] I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), 2017, pp. 542-546, doi: 10.1109/CIAPP.2017.8167276.

[10] P. V. Rajeev and V. S. Rekha, "Recommending products to customers using opinion mining of online product

reviews and features," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], 2015, pp. 1-5, doi: 10.1109/ICCPCT.2015.7159433.

[11] Nallapati, R., Zhou, B., Gulcehre, C. and Xiang, B. (2016) 'Abstractive text summarization using sequence-to-sequence rnns and beyond', 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL). Berlin, Germany, August 2016, pp. 280-290 doi:10.18653/v1/k16-1028

[12] Zhou, Q., Yang, N., Wei, F., Huang, S., Zhou, M. and Zhao, T. (2018) 'Neural document summarization by jointly learning to score and select sentences,' International Conference on Acoustics, Speech and Signal Processing (ICASSP), 18(9), Melbourne, Australia, pp. 59-110, doi: 10.18653/v1/p18-1061

[13] Yu, S., Su, J., Li, P. and Wang, H. (2016) 'Towards high performance text mining: a TextRankbased method for automatic text summarization', International Journal of Grid and HighPerformance Computing (IJGHPC), 8(2), pp.58-75, doi: 10.4018/IJGHPC.2016040104

[14] Rush, A.M., Chopra, S. and Weston, J. (2015) 'A neural attention model for abstractive sentence summarization', Conference on Empirical Methods in Natural Language Processing (EMNLP). Lisbon, Portuga, September 2015, pp. 379-389

[15] See, A., Liu, P.J. and Manning, C.D. (2017) 'Get to the point: Summarization with pointergenerator networks', 55th Annual Meeting of the Association for Computational Linguistics, ACL. Vancouver, Canada, July 2017, pp. 1073-1083, doi: 10.18653/v1/P17-1099

[16] Khatri, C., Singh, G. and Parikh, N. (2018) 'Abstractive and extractive text summarization using document context vector and recurrent neural networks', the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data. London, United Kingdom, August 2018, pp. 1150-1157.

[17]O. Tas and F. Kiyani , "A SURVEY AUTOMATIC TEXT SUMMARIZATION", PressAcademia Procedia, vol. 5, no. 1, pp. 205-213, Jun. 2017, doi:10.17261/Pressacademia.2017.591

[18] Gaikwad, D. K., & Namrata Mahender, C. (2016 March). A review paper on text summarization. International Journal of Advanced Research in Computer and Communication Engineering, 5(3).

[19] George, A., & Hanumanthappa. Document summarization using sentence based topic modelling and clustering. International Journal of Advanced Research (IJAR).

[20] Kherwa, Pooja and Bansal, Poonam (2020) Topic Modeling: A Comprehensive Review. EAI Endorsed Transactions on Scalable Information Systems, 7 (24): e2. ISSN 2032-9407

[21] Bagalkotkar, A., Khandelwal, A., Pandey, S., Sowmya Kamath, S. A novel technique for efficient text document summarization as a service. In 2013 Third International Conference on Advances in Computing and Communications.

[22] Kaikhah, K. (2004 June). Automatic text summarization with neural networks. In Second IEEE International Conference on Intelligent Systems.

[23] Usmani, Usman & Haron, Nazleeni & Jaafar, Jaafreezal. (2021). A Natural Language Processing Approach to Mine Online Reviews Using Topic Modelling. 10.1007/978-3-030-76776-1_6.

[24] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.

[25] Nallapati, R., Zhai, F., Zhou, B.: Summarunner: A recurrent neural network based sequence model for extractive summarization of documents, The Thirty-First AAAI Conference on Artificial Intelligence (2016).

[26] Steyvers, M., and Griffiths, T. (2007). ―Probabilistic topic models‖. In T. Landauer, D McNamara, S. Dennis, and W. Kintsch (eds), Latent Semantic Analysis: A Road to Meaning. Laurence Erlbaum

[27] Anvitha Aravinda, Gururaja H S, Padmanabha J, 2022, Unique Combinations of LSTM for Text Summarization, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICEI – 2022 (Volume 10 – Issue 11),

[28] Rajendran, G.B., Kumarasamy, U.M., Zarro, C., Divakarachari, P.B. and Ullo, S.L., 2020. Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images. Remote Sensing, 12(24), p.4135.

[29] TK. Shashank, N. Hitesh, HS. Gururaja, Application of Few-Shot Object Detection in Robotic Perception, Global Transitions Proceedings,2022,ISSN-2666- 285X, https://doi.org/10.1016/j.gltp.2022.04.024.

[30] R Kotadiya, S Bhatt, U Chauhan, Advancement of Text Summarization Us-ing Machine Learning and Deep Learning: A Review, Proceedings of First International Conference on Computing, Communications, and CyberSecurity, 2020.

[31] Shrenikaa, S., GouriPriyaRamini, L. and Geetavani, B., 2021. Abstractive Text Summarization By Using Deep Learning Models. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(14), pp.2404-2411.

[32] Alghamdi, Rubayyi & Alfalqi, Khalid. (2015). A Survey of Topic Modeling in Text Mining. International Journal of Advanced Computer Science and Applications. 6. 10.14569/IJACSA.2015.060121.

[33] Blei, D.M., and Lafferty, J. D. ―Dynamic Topic Models‖, Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, 2006.

[34] Hofmann, T., ―Unsupervised learning by probabilistic latent semantic analysis‖, Machine Learning, 42 (1), 2001, 177- 196.

[35] Kakkonen, T., Myller, N., Sutinen, E., and Timonen, J., ―Comparison of Dimension Reduction Methods for Automated Essay Grading‖, Educational Technology &Society, 11 (3), 2008, 275-288.

[36] Liu, S., Xia, C., and Jiang, X., ―Efficient Probabilistic Latent Semantic Analysis with Sparsity Control‖, IEEE International Conference on Data Mining, 2010, 905-910.

[37] Bassiou, N., and Kotropoulos C. ―RPLSA: A novel updating scheme for Probabilistic Latent Semantic Analysis‖, Department of Informatics, Aristotle University of Thessaloniki, Box 451 Thessaloniki 541 24, Greece Received 14 April 2010.

[38] Romberg, S., Hörster, E., and Lienhart, R., ―Multimodal pLSA on visual features and tags‖, The Institute of Electrical and Electronics Engineers Inc., 2009, 414-417.

[39] Hofmann, T. (1999) Probabilistic latent semantic analysis, In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, pp.289_296.Morgan Kaufmann Publishers Inc.

[40] Blei, D.M., Ng, A.Y., and Jordan, M.I., ―Latent Dirichlet Allocation‖, Journal of Machine Learning Research, 3, 2003, 993-1022.

[41] Zhi-Yong Shen,Z.Y., Sun,J., and Yi-Dong Shen,Y.D., ―Collective Latent Dirichlet Allocation‖, Eighth IEEE International Conference on Data Mining, pages 1019–1025, 2008.

[42] X. Wang and A. McCallum. ―Topics over time: a non-markov continuous-time model of topical trends‖. In International conference on Knowledge discovery and data mining, pages 424–433, 2006.

[43] J. L. Neto, A. A. Freitas, and C. A. Kaestner, "Automatic text summarization using a machine learning approach," in Advances in Artificial Intelligence. Springer, 2002, pp. 205-215.

[44] K. M. Svore, L. Vanderwende, and C. J. Burges, "Enhancing singledocument summarization by combining ranknet and third-party sources." in EMNLP-CoNLL, 2007, pp. 448-457.

[45] D. Hingu, D. Shah, and S. S. Udmale, "Automatic text summarization of wikipedia articles," in Communication, Information & Computing Technology (ICCICT), 2015 International Conference on. IEEE,2015, pp.I-4.

[46] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields." in IJCAI, vol. 7, 2007, pp. 2862-2867.

[47] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001, pp. 19-25.

[48] R. Mihalcea, "Language independent extractive summarization," in Proceedings of the ACL 2005 on Interactive poster and demonstration sessions. Association for Computational Linguistics, 2005, pp. 49-52.

[49] N. Lalithamani, R. Sukumaran, K. Alagamrnai, K. K. Sowmya, V. Divyalakshmi, and S. Shanmugapriya, "A mixed-initiative approach for summarizing discussions coupled with sentimental analysis," in Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing. ACM, 2014, p. 5.

[50] Yookyung Jo, John E. Hopcroft, and Carl Lagoze. The Web of Topics: Discovering the Topology of Topic Evolution in a Corpus, The 20th International World Wide Web Conference, 2011.