

Summary in General Summary of an Overview of Opinion Mining

Rifat Mustafa , Shabana Rai, Ubaid Ullah, Muhammad Sohaib Naz

¹#Faculty of Computing The superior College Lahore

²#Faculty of Computing ,University of Lahore

³# Faculty of Computing ,Govt university Faisalababd (GCUF)

⁴# Faculty of Computing University of Agriculture Faisalabad

Email: rifatmustafa9214@gmail.com , Shabanarai85@gmail.com, ubaid.aleem@riphahfsd.edu.pk, m.sohaib@riphahfsd.edu.pk

ABSTRACT

The internet is a very effective resource for solving all problems in the present era. The world's population as a whole spends one-third of their time and money using the internet. People learn things from it in every aspect of life, including education, entertainment, communication, shopping, etc. In order to achieve this, consumers take use of websites and share comments or opinions about various goods, services, events, etc. based on their personal experiences. In this way, the input from those webs is composed into a sizable amount of textual data that can be investigated, assessed, and controlled for the decision-making process. Natural Language Processing (NLP) and the extraction of the key theme are both aspects of opinion mining (OM). It's crucial to get feedback on our ideas from people in the form of good, negative, and neutral remarks. Researchers are now presenting information in the form of summaries that will benefit a variety of people. Since the 1950s, the research community has generated automatic summaries, but there are two distinct types of automation: abstractive and extractive approaches. A brief summary of opinion mining's automation process approaches and summarization techniques is provided in this paper.

ARTICLE INFO

Article History:

Received 20/8/2022

Revised 19/11/2022

Accepted 13/4/2023

Available,online
15/4/2023

Keywords:

Natural Language

Sentiment Analysis

Opinion Mining

Text summarization

1.INTRODUCTION

In the field of computer science, the idea of data mining is becoming more and more well-known. Mathematical methods are those that incorporate quantitative techniques. Some of the mathematical equations, algorithms, prime methodologies,

conventional logistical regression, and neural networks may be among these mathematical techniques. These are all mathematically based methods [1]. The process of categorising a vast amount of data is known as data mining. According to data analysis, it is to categorise trends and resolve data issues [2]. Analysis is a technique where a statement is broken down into smaller parts

and lessons are learned from each one about how each part functions and interacts with the other [3]. A study of how people behave in relation to their textual content Opposition Mining (OM). It is a sort of natural language processing (NLP) that is used to examine public sentiment towards a specific good or service [4, 5]. OM, also known as Sentiment Analysis (SA), entails creating a system to gather and organise opinions regarding a good or service. SA is the forerunner of OM, which studies how people feel about a particular subject. (positive, negative, and neutral). The goal of SA is to ascertain a subject's attitude towards a topic, the overall textual polarity, or their emotional response to a piece of writing, an interaction, or an event. The goal of OM is to produce opinions from facts in text form. Numerous organisations invest their funds. Includes tools for sentiment analysis and opinion research [6]. Similar to this, everyone—including organisations and individuals—wants to know what the public thinks about a certain good or service, issue, occasion, etc. This form of survey is useful for research. Therefore, data collecting is made simpler by the variety of sources that are available, including blogs, web forums, discussion platforms, comment boxes, etc. The main focus of research is the extraction of information and knowledge. Because web data is changing, it can be challenging to retrieve this type of information [7].

2. RECENT WORK

This section gives a quick review of OM and the work of several scholars. "Given a set of evaluative text documents D that contain opinions about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document $d \in D$ and to determine whether the comments are positive, negative, or neutral," [8] says Liu of the OM issue. Figure 1 depicts the four issue categories that OM maintains.

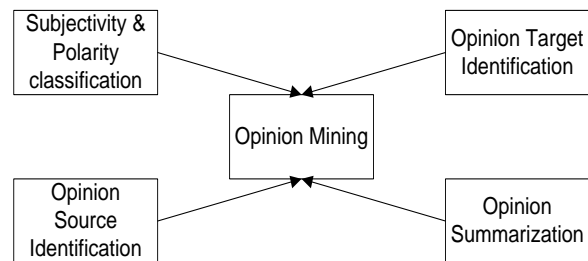


Fig.1 Opinion mining tasks

Previously, a numerous OM techniques proposed by different researchers with multilingual based work. The focus of this research is evaluation process about text classification performance in which examines the algorithms performance on different datasets [9]. Figure 2 shows the steps of the evaluation process.

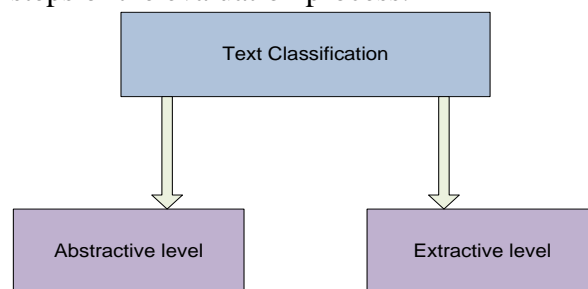


Fig.2 .Evaluation Process

Abstractive level NLP creates an abstract summary of the text while maintaining the initial conceptual knowledge of any form of data during the evaluation process. On the other hand, extractive level uses learning methods that are used to train several classifiers to determine the relevance between the chosen sentences and the source texts [10]. We start with our area of interest for OM research and methodology. We will now go over the summary of user feedback on the documents. The main goal of this summary is to draw concepts from the reviews and align them with the resulting opinion trajectories. The dimensions are worked on by intensity features as they extract the entities. Dimensions provide summary through ranking with the aid of opinion strength [11–12].

3. RESULTS AND DISCUSSIONS

The sentences and words in the source document are the main extraction concern for statistical approaches. These methods are independent, so they don't need any extra or different language-related statistical data, such as sentence location, optimistic and non-optimistic reserved words, sentence similarity, comparative length, numerical data, the proper noun in the sentence, sentences thick pathway, aggregated similarity, etc. Following this, a high-acquire calculated sentence is assisting in producing a decent summary [13]. Figure 3 depicts the many stages of statistical methods used for extracting text summaries.

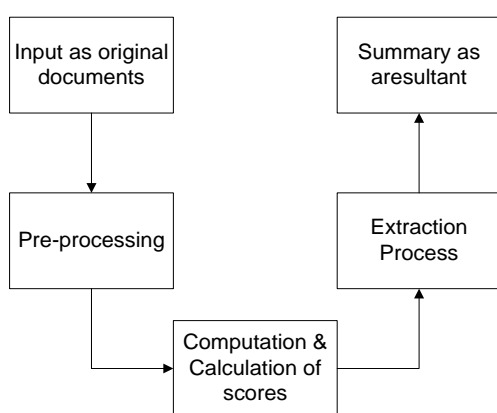


Fig.3 Automatic extractive process by using statistical methods

The topical region of the document under discussion is known as the topic, and it contains all of the written information that describes the manuscript. The main concept pertaining to procedures that repeatedly arise is the establishment of topic design [14]. Five sections make to the representation of the topic, as seen in Figure 4.

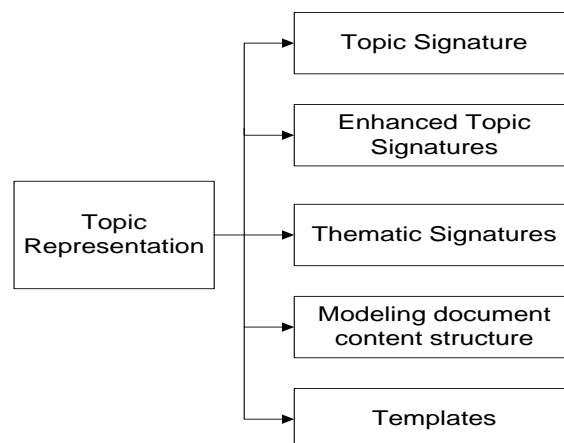


Fig.4 Classification of topic representation

These methods provide phrases or words as nodes while representing semantic representation as edges. For the extractive process of text summarization, many levels of graphical approaches are displayed in Figure 5.

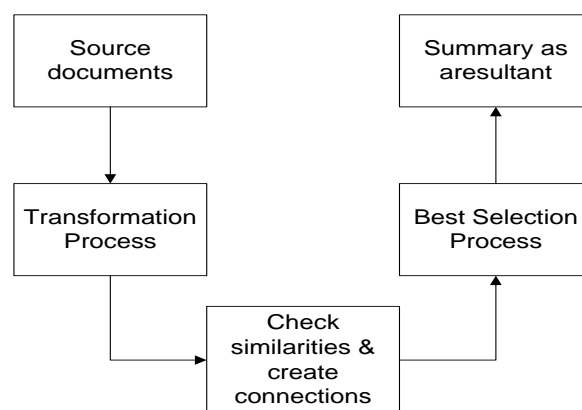


Fig.5 Automatic extractive process by using graphical methods

With this technique, connections between words or sentences are determined. Mann and Thompson provide a Rhetorical Structure Theory (RST) in the subject of estimating in linguistics, which generates two fundamental criteria for processing. The first criterion concerns consistent text, which retains minor text components and these components related with links. To assess the effectiveness of the structure, the second rule inspection of the links is performed [15–

16]. The fundamental principles of Rhetorical Structure Theory (RST) are shown in Figure 6 below.

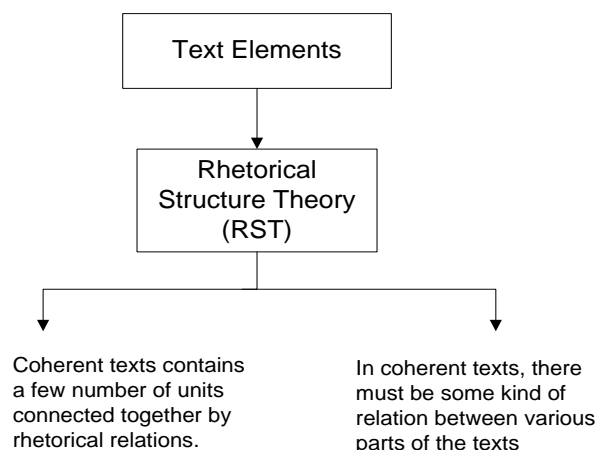


Fig.6 Rhetorical Structure Theory in discourse method

Using supervised, unsupervised, and semi-supervised approaches, three major categories of machine learning are used to track words. Regression [17], Decision Trees (DTs), Multi Layer Perceptrons (MLP), and then they extended their work in accordance with the Support Vector Machine (SVM) and Navie Bayes (NB) classifiers for this purpose [18] were all employed by Fattah and Ren under these categories. Information exchange between physical and virtual machines.

4. CONCLUSION

Text summarization is currently a relatively well-known area of study. The major goal of this study is to summarise earlier, significant work in the field, current advancements, and future directions. The author of this paper provides an overview of extractive and abstractive text summarising methodologies as well as recent developments in these methods, making this research a great place to start for aspiring new researchers. Tables 3 and 4 provide a detailed comparison of the baseline procedures, datasets, calculated methods, and measurements used in the categories of extractive and abstractive text summarization. According to their methods, text summarization requires some additional improvements, and several

major issues have emerged in this area of research, such as new characteristics with semantics related to words and sentences, modifications in the categories of the creation of summaries, the necessity for additional memory-capable processors in linguistic approaches, modifications to coherent summary contents, and improvements in the procedures used to evaluate summaries the creation of summaries, the necessity for additional memory-capable processors in linguistic approaches, modifications to coherent summary contents, and improvements in the procedures used to evaluate summaries.

REFERENCES

- [1] Li, P., Bing, L., Lam, W., Li, H., & Liao, Y. (2015b, April). Reader-Aware Multi-Document Summarization via Sparse Coding. In IJCAI (pp. 1270-1276).
- [2] Parveen, D., & Strube, M. (2015, July). Integrating Importance, Non-Redundancy and Coherence in Graph-Based Extractive Summarization. In IJCAI (pp. 1298-1304).
- [3] Yao, J. G., Wan, X., & Xiao, J. (2015, June). Compressive Document Summarization via Sparse Optimization. In IJCAI (pp. 1376-1382).
- [4] Banerjee, S., Mitra, P., & Sugiyama, K. (2015, July). Multi-Document Abstractive Summarization Using ILP Based Multi-Sentence Compression. In IJCAI (pp. 1208-1214).
- [5] Oufaida, H., Blache, P., & Nouali, O. (2015, June). Using Distributed Word Representations and mRMR Discriminant Analysis for Multilingual Text Summarization. In International Conference on Applications of Natural Language to Information Systems (pp. 51-63). Springer, Cham.
- [6] Gupta, V. (2013). Hybrid algorithm for multilingual summarization of Hindi and Punjabi documents.

- In Mining Intelligence and Knowledge Exploration (pp. 717-727). Springer, Cham.
- [7] Bairi, R., Iyer, R. K., Ramakrishnan, G., & Bilmes, J. A. (2015). Summarization of Multi-Document Topic Hierarchies using Submodular Mixtures. In *ACL (1)* (pp. 553-563).
- [8] Kedzie, C., McKeown, K., & Diaz, F. (2015). Predicting Salient Updates for Disaster Summarization. In *ACL (1)* (pp. 1608-1617).
- [9] Hong, K., Marcus, M., & Nenkova, A. (2015). System Combination for Multi-document Summarization. In *EMNLP* (pp. 107-117).
- [10] Yao, J. G., Wan, X., & Xiao, J. (2015). Phrase-based Compressive Cross-Language Summarization. In *EMNLP* (pp. 118-127).
- [11] Khan, A., Salim, N., & Kumar, Y. J. (2015). A framework for multi-document abstractive summarization based on semantic role labelling. *Applied Soft Computing*, 30, 737-747.
- [12] Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.
- [13] Ouham, S., Hadi, Y., & Ullah, A. (2021). An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model. *Neural Computing and Applications*, 33(16), 10043-10055.
- [14] Ullah, A., & Nawi, N. M. (2021). An improved in tasks allocation system for virtual machines in cloud computing using HBAC algorithm. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.
- [15] Ouham, S., Hadi, Y., & Ullah, A. (2021). An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model. *Neural Computing and Applications*, 33(16), 10043-10055.
- [16] Sebai, D., & Shah, A. U. (2022). Semantic-oriented learning-based image compression by Only-Train-Once quantized autoencoders. *Signal, Image and Video Processing*, 1-9.
- [17] Alam, T., Ullah, A., & Benaida, M. (2022). Deep reinforcement learning approach for computation offloading in blockchain-enabled communications systems. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.
- [18] Bing, L., Li, P., Liao, Y., Lam, W., Guo, W., & Passonneau, R. J. (2015). Abstractive multi-document summarization via phrase selection and merging. *arXiv preprint arXiv:1506.01597*.