

Analysis of Effective Approaches for Legal Texts Summarization Using Deep Learning

Ms. Shruti J. Sapra (Thakur)¹, Dr. Avinash S. Kapse²

¹Research Scholar, Department of Computer Science and Engineering, Amravati University,

Assistant Professor, Department of CSE, G H Rasoni College of Engineering, Nagpur, Maharashtra, India

²HOD, Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

ABSTRACT

Now a day there is a great importance is associated with the text summarization as there is a large amount of data is available on the different social media platforms, websites and blogs. Also there are large number of tools used for the summarization of the text available in the different forms this text which is available from the different sources are present in the various format some of the data is in the structured and some are in the unstructured format. The user or the application which is using this data has to be careful about the use of the data as the data is available has to be authentic and must be delivered to the user in the proper format. The large amount of time is associated with understanding the meaning of the data. Therefore user has to invest large amount of time in reading the data in other words there is a great deal in getting the meaning of the data as the data is in large amount and hence getting the meaning from such a huge amount of data is not a simple job. The best option is to summarize means to shorten the data in so that user will require less amount of time in understanding the meaning of the data and will be useful for the user or the application in which it is used. This paper essentially discusses the various approaches from the domain of the deep learning and machine learning in shortening the data as well as summarizing the data. Different algorithms are also discussed in order to understand the various parameters associated with the text summarization. Also the best algorithm must be chosen for the effective text summarization process for the better results and better efficiency. Also in India there is a great importance is associated with the legal text but many a times the common man or the user is unable to recognize the meaning of the text by means of the text is very large and descriptive hence it is very essential to decode such a text in simple and short text. Thus the specialized algorithms are used to interpret the meaning or directly the text in to the format the use is able to understand. In other words the techniques have been discussed shortens or summarizes the legal text and makes easier or readable for the application or the user.

Keywords: Deep learning algorithms, legal text summarization, Abstractive text summarization, strategies for summarization.

I. INTRODUCTION

Text summarization is the restating of the actual information of the text and make that text as short as

possible or in other words expresses the large information in a very few words or sentences as it is possible. [1] This process will significantly reduce the size of the text and but the information should

convey the same meaning as previously conveyed. This increases the readability of the text for the user of the application. [2] And this task can be done using the various tool such as artificial intelligence machine learning also with the support of dissimilar deep learning algorithms available in the previous research [3]

Text summarization also referred as the unique the method of shortening very long words, sentences or paragraphs of text or data. The clear intention behind this is to make coherent and short summary having the important notes which is short text or data which is used in number of applications for the purpose of proper understanding in small amount of time. [4]

Many times the text summarization is Automatic in nature and is a major issue in medical field and hence different techniques has to be applied for such a type of data which is very crucial. [4]

Summarization can also be defined as reduction in the major content of the information whereas the original meaning of the information. Has to be there in the short text .which will be conveyed to the intended user of the information.

Most of the author has worked on the Abstractive summarization method which is the unique method of summarization of the text and essentially very useful considering the different domain of the application and has number of advantages over the other techniques discussed [5], On the other hand is a method where the short summary is produced by creating new paragraphs by either replacing or reconstructing the paragraphs, instead of making use of different use of words having the same meaning and the important paragraphs.[6]

Legal text can be various kinds of information written for different purposes related to the law, including: book about law. Legal treatise, Law book, a major publications related to all the law or related to field of law. In other words, texts created in the course of legal research. [7]

II. BACKGROUND

Different researchers from the different areas studied different aspects of the text summarization specially the text summarization in concerning with the legal text. They found the various parameters which affect this process of summarization and also used different methods for the text summarization.

Summarization of Text is the complete process of filtering the significant information from different sources like internet and others to create a unique version for a particular applications or user and work. [7]

There are mainly two types of summarization of the text in this regard which are discussed as below

a. Abstractive Summarization

Abstractive summarizers are one of the efficient methods of summarization in which sentences are not selected from the main given information or passage to create the summary. Instead, they may create a paraphrasing of the related main information of the given information, using a different set of vocabulary from the main information.

b. Extractive summarization

Extractive summarization means identifying significant sections of the information and producing then producing a unique subset of the paragraphs from the main information; while abstractive summarization may produce matter in a completely new way after completion of this process of summarization of the information using various natural language processing tools

From the important findings from the previous research it is found that the abstractive summarizer performs more better over all the different methods of the summarization in concerning with the different parameters of the

summarization, the outcomes recommend that the margin by which abstraction outperforms extraction is greater when controversialist is high, providing a context in which the need for generation- based methods is especially great.

There is another approach which is used in the text summarization i.e. Latent Semantic Analysis which is an unsupervised learning algorithm and thus can be applied for extractive summarization of text. First, import the summarizer. [7]

The second approach used in the process of summarization is the application of different tools. These tools are used according to the application need. Some of the applications requires the basic summarization while the others requires high level of the summarization of the text. The advanced level of summarization requires the high level of tools which are very efficient in summarization and also requires very less amount of time for the complete process of summarization on the other hand the basic tools are not that much efficient as compared to the advanced tools but are more cost efficient as compared to the advanced tools. Thus the user will select according to their preferences which tool is too used for the summarization. In case of basic tools there may be an error in summarizing the text or information and May sometimes leads to wrong information in the summarized text. [8]

Therefore it is always a better choice to use the advanced tools for thee summarization. These tools involve the application of advanced algorithms hence they are more effective as compared to the basic algorithms. These advanced algorithms are the nothing but deep learning algorithms also advanced machine learning algorithms.

The paper is arranged in following manner:

Section I Introduction. Section II describes Background. After that Section III describes previous work done by various researchers in this domain.

Section IV describes various existing methodologies. Section V describes various attributes and different parameters. Section VI gives new approach. Section VII experimental tests carried out. Section VIII is approaches for Testing's, Section IX is outcome and results. Section X is conclusion. Finally Section XI is future scope of this analytical paper.

III. PREVIOUS WORK DONE

Some of the approaches use the advanced machine learning and neural network algorithms which are more efficient in summarizing the text or information. The main advantages of such algorithm are they require less amount of time for the process of summarization and mainly convey the same information as the main or original text was conveying. One of the best approaches is the application of Bidirectional transformer is a best transformer applied to overcome the limitations of RNN and other neural networks as defining characteristics of Long term dependencies present in the information or the text. It is a pre-trained and well defined structure which is very efficient and is naturally bidirectional. This well defined structured approach can be tuned to easily to perform various functions of NLP tasks as specified in the different application or as per the requirement of the user which will suit the requirement of the user for the particular application.

Hence this model or structure is more likely to use in most of the summarization process and is very beneficial for the particular application point of view. There is another approach which is used by most of the researchers in the summarizing the large amount of text and i.e. PEGASUS this is Advanced Model or structure for Abstractive and extractive Summarization of text and is also the best approach for using the different sentences or in other words replacing the same meaning sentences by another sentences or paragraphs this approach will also

significantly reduce the text to a short text and is also a better choice for the summarization of the text or information.

This approach for summarization of text is one of the most challenging and difficult task in case of legal text is involved as there is a use of very critical rather a very specific words are used in this context an many times the another word or sentence for the same word or sentence is not suitable for the replacement and in such a case the tasks can be completed with the help of specific algorithms in processing a natural language which may involve understanding of long paragraphs in legal context and find the meaning of the each sentence or words in the paragraph. Sometimes this method also use the approach known as information or data compression and producing a new language, this approach is also a typical approach and also involves many challenges which has to be solved using a specific algorithms and special techniques.[8]

IV. EXISTING METHODOLOGIES

There are different methodologies which are efficient in summarizing the text some are very advantageous and applicable in the wide range of the applications hence they are more popular in this context and also are generating effective results as compared to the other methods of summarization [8] These methods and techniques may consider different or same parameters for the summarization like accuracy performance and time required for the performance. Researchers deeply understand the phenomenon of summarization and then applied the different methods and techniques according to the application domain. [9]

Out of which some of the methods are discussed here some of the methods explained [3] gives importance to the repetitive words, sentences and paragraphs used in the summarization and also used for the specific application. Their main aim is to reduce the

size of the text or information while the meaning should not change accordingly. Therefore they Include keywords and terms used by the different researcher and writers and carefully think over the text or information and also about how the sources of different ideas which are relevant to the application domain and must be suitable for the application where they are going to use. Applying only short notes, might explain the genuine writer's main concepts in different form. Then explain how those concepts for the same context.

The next approach which is used in the Google is also studied here and also helpful in understanding the minute details of the summarization of the text which is also very useful in many context. This approach explain the importance of the Not to be outdone for the particular text or information which is used in the legal context and therefore be applied for the legal text or information summarization, a Google Brain and his team in the Imperial College London has developed a complete system and such Pre-training with more and efficient extracted text in creating the Gap-sentences for Summarization of typical text like legal text and legal information like court judgments and court amendments [10] The clear sentence to sentence sequence is applied to maintain the flow of information and make the meaning clear to the user or the application

This approach also used the concepts of Pegasus in which the summarization of the legal text and then interpret in the different form is done and in which the user can understand and interpret this shot text.

This approach is also applicable for leverages Google's Transformers structure which may be combined with predefining the various objectives tailored according short text production.

V. ANALYSIS AND DISCUSSION

Analysis of the summarization process is based on the different parameters such as word count, number of

sentences reduced and number of paragraphs reduced. In addition to this there are various parameters relating to the technology are the time required for the summarization, space required for the text or information before and after the summarization of the text or information. Then there are parameters relating to the application in which these methods or techniques applied. All the approaches are studied in detail and analyzed according to the different parameters discussed above.

In addition to this there are more important parameter is the used to technology and the use of the algorithm using which the summarization of the text is done.

VI. PROPOSED METHODOLOGY

This method proposed in this section is very useful for the documents relating to the court judgments and applied in various applications relating to the legal aspect. The different head notes and sections of a various court Judgments regarding the legal document summarization use the case of text summarization but it does not relate to the paragraphs of the other judgments of the courts of the other nation.

This method comprises of two main steps, first is the Generation of better quality information by expressing the short notes present as headings of the documents of the legal notes and second is applying this information in context to extraction of the significant words or paragraphs mentioned in the document which is to be inserted. This research proposed has several benefits over the method discussed earlier. For the implementation of the algorithm. The word count table is generated for the words those are repetitive in nature in the given legal text such words are replaced by the same meaning words from the dictionary of the dataset and a new short summary is created which is very short in nature as compared to the previous original text and

that also can be used in the different applications like souckal media, blogs and other websites where summarization is essential. Thus this approach maintains a special dictionary for the words in the table which are recitative and which are having the same meaning as that of the other.

After this process of word replacement sentences are Tokenize the replacement and similar or repetitive meaning sentences are replaced using a dataset. Thus the set of sentences are replaced in a sequence to form a short notes or short text. This process is continuously implemented to generate a summary which can be directly applied to the application of domain.

VII. APPROACH FOR TESTINGS

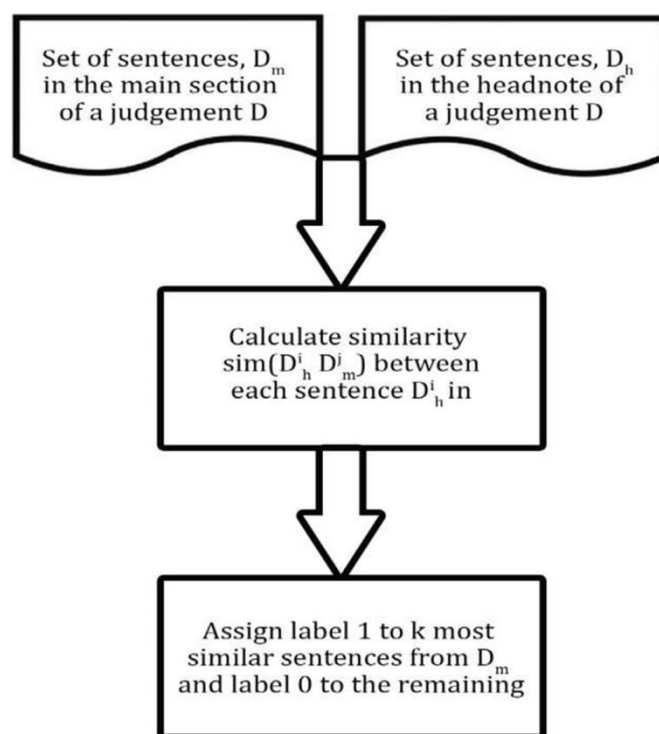


Fig 1. : Patterns and data set generation in the legal text.

Labelled dataset generation

This stage involves the production of various useful dataset for the replacement of the data from the actual information. Thus using this ready data set the

short text or information is produced and applied to the different applications. Some of the case judgments may contain short summary which is not to be replaced and must match with our dataset and this will significantly save the time for the summarization. Such an information is highlighted as not to replace. Similarly, there are some words or significant sentences which are not to be replaced such words and sentences have again to be labeled for non replacement. This approach used the Text Rank which is a graph-based ranking structure. Text processing can be done in order to find the related paragraphs or related sentences in text and also to find useful words and after that replaced them continuously

Single-document is summarized using transformation of the source text into a compressed, shorter text in which the irrelevant data is deleted. Using this approach large and complex legal documents are summarized in the best short form which are easily readable and also takes less space on the disk.

VIII. IMPLEMENTATION

The implementation of the proposed method is done using a deep learning algorithm which are very efficient and follows the advanced text summarization approach in which less amount of time is needed and following procedure is applied

The proposed approach can be tested using the different tools of the summarization and results are generated which are compared against the different essential parameters and best summary is generated.

IX. OUTCOME AND RESULTS

It is very clear from the above proposed method that the different parameters studied according to different methods behave differently in every situation and hence are useful for this approach. This proposed method is very useful in all

scenarios and various applications where it is implemented and hence found suitable for the real time and other applications also.

X. CONCLUSION

This contribution is a multidimensional approach in which this proposed approach is a new technique of generating a unique summary which is the need for expert summarization. This approach finds the best in such a complex domain in which other techniques of summarization may not work efficiently. This approach finds the most suitable technique in the field of summarization of the text or information from the legal documents.

XI. FUTURE SCOPES

It is expected that the continuous research and development will eventually result in a variety of utility and tools. These strategies will also improve the effectiveness and efficiency for implementing various methods and technologies of text summarization of legal information. So, overall enhanced summarization approach will improve the proposed method significantly.

XII. REFERENCES

- [1]. A full Python Implementation of the ROUGE Metric (not a wrapper). Retrieved from <https://github.com/pltrdy/rouge/>.
- [2]. Alguliev, R., Aliguliyev, R., 2007. Experimental investigating the F-measure as similarity measure for automatic text summarization. Appl. Comput. Math. 6 (2), 278–287.
- [3]. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., Kochut, K., 2017. Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.

- [4]. Chen, Y.C., Bansal, M., 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. arXiv preprint arXiv:1805.11080.
- [5]. Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A., 2017. Supervised learning of universal sentence representations from natural language inference data. arXiv preprint arXiv:1705.02364.
- [6]. General purpose unsupervised sentence representations. Retrieved from <https://github.com/epfml/sent2vec>.
- [7]. Infsent Sentence Embedding. Retrieved from <https://github.com/facebook research/InferSent>
- [8]. Kanapala, A., Pal, S., Pamula, R., 2019. Text summarization from legal documents: a survey. *Artif. Intell. Rev.* 51 (3), 371–402.
- [9]. Kasture, N.R., Yargal, N., Singh, N.N., Kulkarni, N., Mathur, V., 2014. A survey on methods of abstractive text summarization. *Int. J. Res. Merg. Sci. Technol* 1 (6), 53–57.
- [10]. Kavila, S.D., Puli, V., Raju, G.P., Bandaru, R., 2013. An automatic legal document summarization and search using hybrid system. In: *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*. Springer, Berlin, Heidelberg, pp. 229–236.
- [11]. Keras: The Python Deep Learning library. Retrieved from <https://keras.io/>. Kim, M.Y., Xu, Y., Goebel, R., 2012. Summarization of legal texts with high cohesion and automatic compression rate. In: *JSAI International Symposium on Artificial Intelligence*. Springer, Berlin, Heidelberg, pp. 190–204.
- [12]. Lin, C.Y., 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*. pp. 74–81.
- [13]. Tom M. Mitchell, Herbert Jones, “Machine learning”, Indian edition. Tata Mc Graw Hill Publications, 1 July 2017.
- [14]. Sukriti Verma, Vagisha Nidhi, "Extractive Summarization using Deep Learning", arXiv:1708.00439v2[cs.CL], 9 Jan 2019.
- [15]. Bhumika, V. Srikanth, "Abstractive Summarization with NLP Makes Easier to Memorize", *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, Volume-8 Issue-5, Jan 2020.