

SEMANTIC SUMMARIZATION OF SENTENCE USING DIFFERENT TECHNOLOGIES

Aniket A. Chavan¹

*Dept. of Computer Science and Engg. Ashokrao Mane Group of Institutions, Kolhapur Maharashtra, India
anurajchavan29@gmail.com*

Mr. Pradip A.Chougule²

*Dept. of Computer Science and Engg. Ashokrao Mane Group of Institutions, Kolhapur Maharashtra, India
Chougule.pradip@gmail.com*

----- *** -----

Abstract:- Data mining is one of the extensive practices that is used for handling enormous amount of data. Data commonly can be of any type which is haphazard and needs formatting and interrelation. Here is when we have concepts like summarization, semantic feature extraction and classification. Summarization systems are nothing but systems that involve sentence extraction and interrelation that can be used in large online analysis systems or online word extraction systems that involve mapping and classify words. sentiment level analysis, polling online news, question-answering systems are few of the applications where summarization is done either at word level or sentence level. Sentence level summarization includes degrading and refining the semantic structure in a way that helps us simplify the summarization.

----- *** -----

I INTRODUCTION

A. Tree and word embedding based sentence similarity for evaluation of good answers in intelligent tutoring system.[1]

Euler calculation was used to create a progression of words dependent on tree and SørensenDice coefficient was applied to decide the comparability between analyzed trees. The accentuation is on characterizing the likeness between the right and wrong answers from the Yahoo Question and Answer of the Non-Factual Data Set. Proposed calculation was utilized on two sorts of trees. First is the electorate tree created by Stanford CoreNLP, and second is hand crafted calculation that produces second sort of tree, called information tree which is gotten from parse tree. In our correlation, Zhuang-Sasha calculation was additionally utilized. Second methodology that was utilized for sentence correlation utilizes Word2Vec model for discovering word inserting's and figuring sentence normal vector, after that cosine distance was applied to decide comparability between two sentences. Results created with this strategy were contrasted and our technique in discovering sentence comparability dependent on information tree. Approach portrayed in this paper can be utilized in assessment of right

answers which will be utilized in our execution of Intelligent Tutoring System.

B. SSE: Semantic Sentence Embedding for learning user interactions.[4]

Semantics is an integral component in NLP. Semantics provides a meaningful view about the meaning of the language. The meaning of the text is susceptible to the negative words present in them. Thus the impact of semantics in NLP in turn increases the impact of negative words in the sentence. It plays crucial role in shaping the semantics of any sentence. Although all the sub-areas of natural language processing try to find out the impact of negative words on their implementation, the effect of these words are not considered in the word embedding process. Being the foremost step in NLP the omission of these words in embedding can affect the representation of the language. The aim is to propose a model which produces word embeddings by considering the negative words. Although most NLP implementing systems consider all negative words this paper focuses on the word 'not' which is followed by an adjective. The antonym for the adjective is identified and is then replaced in the corpus. The model is achieved by replacing

the corpus where negative words and this corpus is fed to a three layer neural network to form the embeddings. The proposed technique proves an accuracy of 93.6% in retrieving the relevant answer to the user.

C. Question Understanding Based on Sentence Embedding on Dialog Systems for Banking Service.[3]

This paper acquaints an inquiry understanding framework with react fitting answers in an exchange framework for banking administrations. The inquiry understanding framework gives an mechanized reaction administration in a particular area (for example banking). This can build reaction pace of a client guiding help, also, improve business effectiveness and skill. The inquiry understanding framework arranges spaces, explicit classifications, and discourse demonstrations of inquiries. At last, the framework investigate implications also, purposes of the inquiries, and looking through right answers even different information sentences. In this paper, we portray techniques for catchphrase tokenizing, design acknowledgment, sentence inserting, investigating discourse expectation, and looking through comparable FAQs. Through these strategies, we have built up the inquiry understanding unit in a genuine intelligent framework for monetary administrations no doubt insurance agencies and banks, and break down the convenience of the framework through commonsense framework usage models.

D. Test-Driven Summarization: Combining Formative Assessment with Teaching Document Summarization.[7]

The dispersion of learning innovations has encouraged the utilization of versatile and Web-based applications to survey the information level of students. In equal, an expanding research premium has been dedicated to concentrating new learning examination instruments ready to sum up the substance of enormous arrangements of learning records. To overcome any barrier between developmental appraisal apparatuses and report outline frameworks, this paper tends to the issue of suggesting short synopses of huge sets of learning archives dependent on the results of multiple choice tests. In particular, it presents another strategy for incorporating developmental appraisal through portable applications also, synopsis of learning records in text based structure. The substance of the different decision tests is misused to drive the age of record rundowns custom-made to explicit subjects. Moreover, the results of the tests are utilized to naturally prescribe the created rundowns to

students dependent on their real necessities. As a contextual investigation, we played out an assessment experience of understudies' advances, which was led in the setting of a college level course. The accomplished outcomes show the appropriateness of the proposed system.

E. Graph based automatic document summarization with different similarity methods.[8]

Today, with the fast expansion in the utilization of the web, a great many assets can be reached about an data that is intrigued. Not withstanding, it is troublesome and time burning-through to figure out which of these sources is helpful. Programmed report rundown is a measurement decrease measure which stays the significant pieces of the content. In this study, the TextRank calculation, which is a diagram based rundown approach, is utilized with 4 distinctive similitude techniques. The impact of these strategies on the consequently produced outlines is analyzed. Among the closeness techniques, Levenshtein strategy was more fruitful than others with 0,506 Rouge-1 score.

F. Single-Document Summarization Using Sentence Embeddings and K-Means Clustering.[6]

This paper proposes a novel technique for extractive single report rundown utilizing K-Means bunching and Sentence Embeddings. Sentence embeddings were prepared by K-Means calculation into various bunches relying upon the required rundown size. Sentences in a given bunch contained comparative data, and the most fitting sentence was picked and remembered for the rundown for each bunch by an edge relapse sentence scoring model. Test ROUGE score assessment of rundowns of different lengths for the DUC 2001 dataset exhibited the viability of the methodology.

G. Multi-document summarization using sentence clustering.[9]

This paper presents a way to deal with inquiry centered multi record rundown [9] by consolidating single archive outline utilizing sentence bunching. Both syntactic and semantic similitude between sentences is utilized for bunching. Single record synopsis is created utilizing report include, sentence reference list highlight, area highlight and idea comparability highlight. Sentences from single record rundowns are grouped and top most sentences from each bunch are utilized for making multi-record rundown. We noticed a normal F-proportion of 0.33774 on DUC 2002 multi-record dataset, which is equivalent to three best performing frameworks detailed on the equivalent dataset.

G. A Context based Coverage Model for Abstractive Document Summarization.[5]

Programmed abstractive synopsis is one of regular language handling fields that creates an arrangement of words speaking to significant data of the info report. The grouping to-succession model, which is generally utilized in abstractive rundown, has a redundancy issue in which a similar sub-design is consistently produced in interpret stages. To take care of this issue, different inclusion models have been proposed in machine interpretation. In programmed rundown, dissimilar to machine interpretation, the lengths of info also, rundown archives are totally different. Since the rundown record is a compacted type of significant importance of the info record. Because of the idea of programmed rundown, it is hard to apply a word position-based inclusion model in machine interpretation straightforwardly. For programmed rundown, we propose a setting based inclusion model to consider the inclusion dependent on the compacted importance of the input record. The setting based inclusion is characterized as the gathered weighted normal of the encoded word significance by the consideration scores. This thinks about the significance of words rather than the situation of words in the information report. In the try different things with CNN/DailyMail dataset, the proposed model shows preferred execution over the past investigations.

II. CONCLUSION

Hence the sentence summarization can be successfully achieved through different technologies that involve use of K-means clustering algorithms, semantic feature extraction, similarity based pattern recognition that can be clubbed with neural networks, machine learning and artificial intelligence techniques.

REFERENCES

1)E. Brajković and D. Vasić, "Tree and word embedding based sentence similarity for evaluation of good answers in intelligent tutoring system," 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2017, pp. 1-5, doi: 10.23919/SOFTCOM.2017.8115592.

2)W. Xing, X. Yuan, L. Li, L. Hu and J. Peng, "Phenotype Extraction Based on Word Embedding to Sentence Embedding Cascaded Approach," in IEEE Transactions on NanoBioscience, vol. 17, no. 3, pp. 172-180, July 2018, doi: 10.1109/TNB.2018.2838137.

3)K. Oh, H. Choi, S. Kwon and S. Park, "Question Understanding Based on Sentence Embedding on Dialog Systems for Banking Service," 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), Kyoto, Japan, 2019, pp. 1-3, doi: 10.1109/BIGCOMP.2019.8679121.

4)J. F. Lilian, K. Sundarakantham, H. Rajashree and S. M. Shalinie, "SSE: Semantic Sentence Embedding for learning user interactions," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944630.

5)H. Kim and S. Lee, "A Context based Coverage Model for Abstractive Document Summarization," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea (South), 2019, pp. 1129-1132, doi: 10.1109/ICTC46691.2019.8939704.

6)S. Agarwal, N. K. Singh and P. Meel, "Single-Document Summarization Using Sentence Embeddings and K-Means Clustering," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 162-165, doi: 10.1109/ICACCCN.2018.8748762.

7)L. Cagliero, L. Farinetti and E. Baralis, "Test-Driven Summarization: Combining Formative Assessment with Teaching Document Summarization," 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, 2017, pp. 401-406, doi: 10.1109/COMPSAC.2017.231.

8)O. Kaynar, Y. E. Işık and Y. Görmez, "Graph based automatic document summarization with different similarity methods," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4, doi: 10.1109/SIU.2017.7960613.

9)V. K. Gupta and T. J. Siddiqui, "Multi-document summarization using sentence clustering," 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), Kharagpur, 2012, pp. 1-5, doi: 10.1109/IHCI.2012.6481826.