Fuzzy Inference-based Models for Extractive Text Summarization

Department of Computer Science and Engineering

- Gagandeep Singh
- Bhavana Mekala
- Nelluri Pavithra Sai Lakshmi
- Santosh Singh Rathore

Overview

- Introduction
- Motivation
- Literature Survey
- Objectives
- Methodology
- Tasks Completed
- Results
- Conclusion
- References

Introduction

- Text summarization has emerged as a useful solution for dealing with large amounts of text, providing users with a concise and consistent summary that captures the essence of the original content.

- Automatic text summarization aims to generate concise summaries that capture the key sentences and relevant information from the original text.

- This explores extractive fuzzy-based text summarizations and their challenges and performance in comparison to other summarization methods to determine their practicality in real-world applications.

- Text summarization using fuzzy rules involves the application of fuzzy logic principles to condense lengthy pieces of text while retaining essential information.

- These rules then guide the process of selecting and prioritizing sentences or phrases in the original text, resulting in a concise and coherent summary that captures the main ideas and key points

Motivation

- Manual text summarization is a time-consuming and costly process
- It becomes impractical when dealing with such massive amounts of text.
- The investigation of automatic text summarizing approaches has been encouraged by the rising demand for effective information digestion.
- it is aimed at reducing the strain of manual summary.
- People frequently experience overwhelming quantities of paperwork due to the spread of the internet and the availability of large amounts of information.
- To guarantee that people can access crucial information without being overloaded with text, it is essential to design efficient and precise text summary systems.
- This has led researchers to explore technological solutions for automatic text summarization .

Literature Survey

AUTHOR(year of publication)	Aim of the Paper	Method used	Results
Adhika Pramita Widyassari [a b], Supriadi Rustad [a], Guruh Fajar Shidik . (2022)	Review of automatic text summarization techniques & and methods	Discussed all the approaches in brief (both abstractive and extractive)	No experiments were done just a compilation of all the studies present.
Wafaa S. El-Kassas [a], Cherif R. Salama [a b] (2021)	Automatic text summarization: A comprehensive survey	Discussed all the approaches in brief (both abstractive and extractive)	No experiments were done just a compilation of all the studies present.

Fábio Bif Goularte [a], Silvia Modesto Nassar [a] (2019)	A text summarization method based on fuzzy rules and applicable to automated assessment	Fuzzy logic	The results show that the proposal provides better f-measure (with 95% CI) than aforementioned methods.
S.A. Babar a, Pallavi D. Patil b(2015)	Improving the Performance of Text Summarization	Fuzzy logic,along with semantic analysis.	The result in the graph shows that our proposed summarizers perform better than fuzzy summarizer approach

Ladda Suanmali1 , Naomie Salim2 and Mohammed Salem Binwahlan3(2009)	Fuzzy Logic Based Method for Improving Text Summarization	Fuzzy logic	The results show the best f-measure for summaries produced by the fuzzy method. Certainly, the experimental result is based on fuzzy logic have improve the quality of summary results that based on the general statistic method.

Dimitrios Novas [a], Dimitrios Papakyriakopoulos [a] (2023)	A ranking model based on user generated content and fuzzy logic	Converts qualitative data to quantitative to perform the ranking using a fuzzy logic.	Achieve their aim to find the best restaurants.
Petr Cintula [1], Petr Hájek [2] (2010)	Triangular norm-based predicate fuzzy logics	The paper surveys the present state of knowledge on t-norm based predicate fuzzy logics	A survey study is conducted

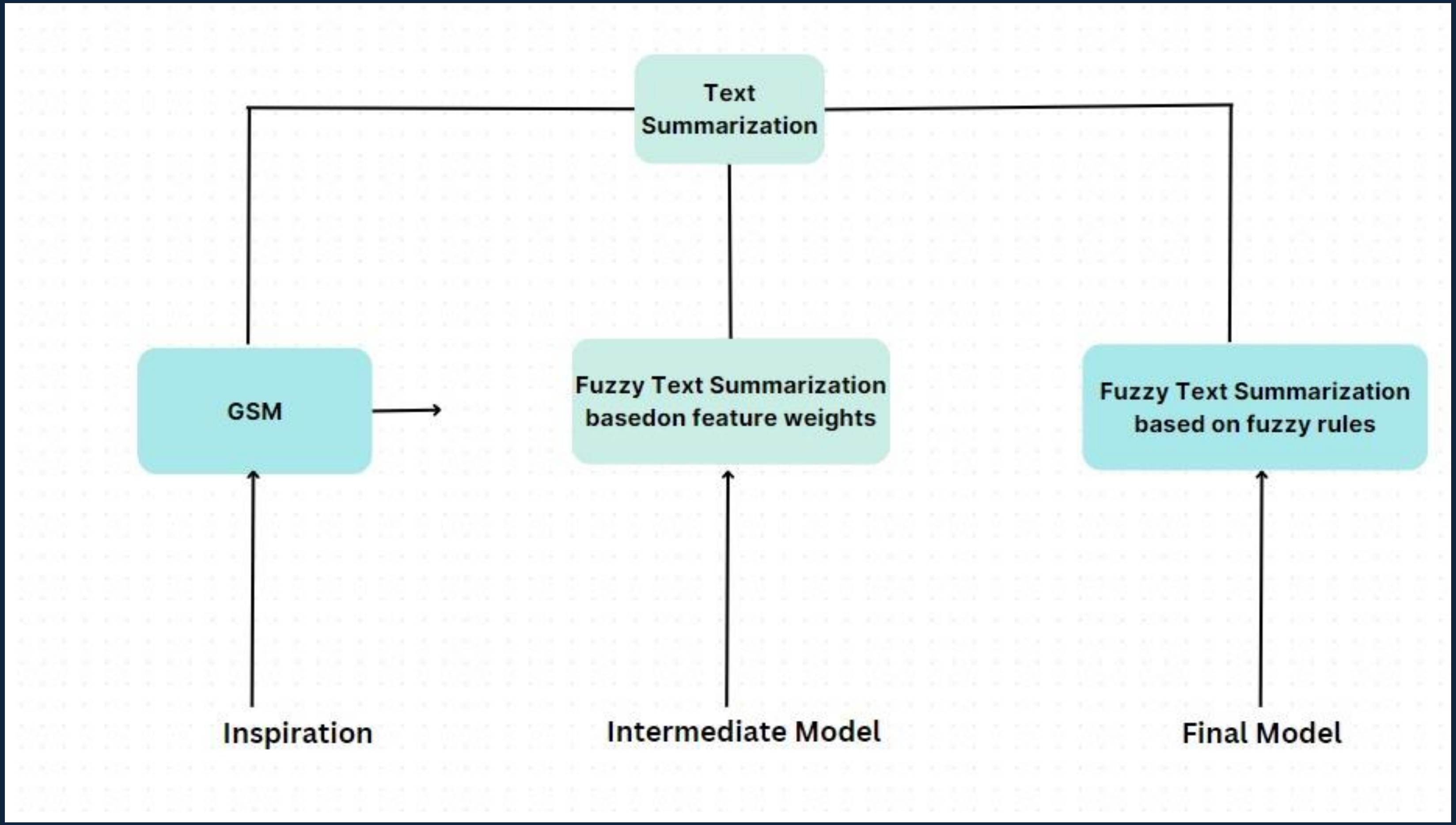
Objectives

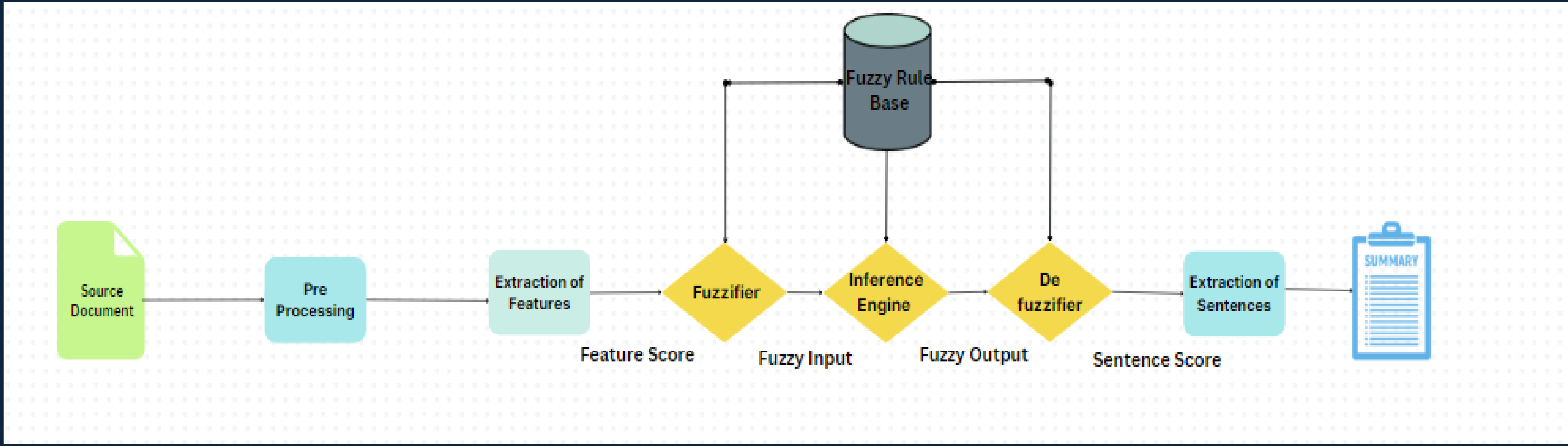
- Implementing Fuzzy-based Model For The Extractive Text summarization

  - Using Fuzzy Rules-in Fuzzy inference

  - Text Summarization using Feature weights by Fuzzy

    - Optimization using Genetic algorithm


Text Summarization
GSM
Fuzzy Text Summarization basedon feature weights
Fuzzy Text Summarization based on fuzzy rules
Inspiration
Intermediate Model
Final Model



Overview of the presented text summarization model

Methodology

Fuzzy-rule-based model for the Extractive Text Summarization:

Generally, text summarization using a fuzzy inference system involves the following necessary steps-

Preprocessing :

1. Text Cleaning
2. Tokenization
3. Stopword Removal
4. Lemmatization or Stemming
5. Part-of-Speech Tagging (POS Tagging)
6. Sentence Segmentation

- After completing the preprocessing steps, every sentence in the document is converted into an attribute vector that contains various features.
- These features serve as attributes aiming to represent the information relevant to their specific purpose.

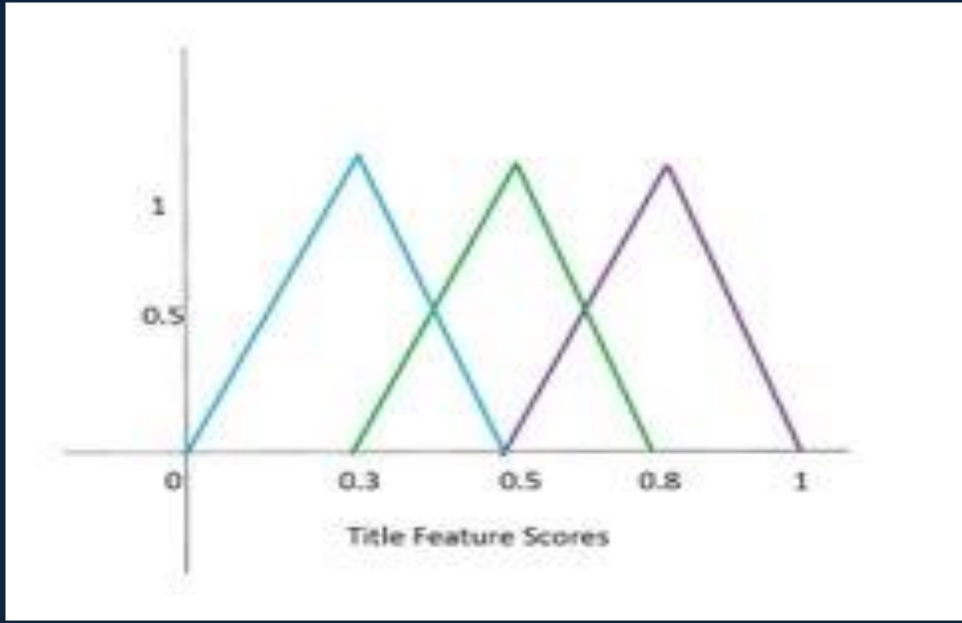
Extraction of feature -

1.Title Feature Score
2.Sentence Length
3.Sentence Position
4.Sentence Similarity
5.Numerical Score Ratio
6.Thematic Word Feature Score
7.Sentence Weight

```
Sentence ratios: [0.38461538461538464, 0.6153846153846154, 0.38461538461538464, 0.6153846153846154, 0.46153846153846156, 0.8461
538461538461, 1.0, 0.6153846153846154, 0.0]
Title feature scores: [0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
Sentence lengths: [6, 9, 6, 9, 7, 12, 16, 9]
Term weights: [0.9999999999999997, 0.9999999999999997, 0.9999999999999997, 0.9999999999999997, 0.9999999999999997, 0.9999999999
999997, 0.9999999999999997, 0.9999999999999997]
Sentence positions: [1.0, 0.875, 0.75, 0.625, 0.5, 0.375, 0.25, 0.125]
Sentence similarities: [0.0, 0.22028815056182974, 0.1908740661302035, 0.474330706497194, 0.4112070550676186, 0.4084036922437499
5, 0.18832393622758892, 0.05224985350000812]
Numerical score ratios: [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.23076923076923078, 0.0]
Thematic word feature scores: [0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
This is outside everything [0.06425179522501522, 0.11706722053599268, 0.08306616995401404, 0.07036607505333747, 0.0892859482502
4673, 0.07267770256164066, 0.07650688628207832, 0.12542916485999214]
```

Model Development:

- **Fuzzification** converts numerical scores into linguistic terms using a triangular membership function .
  Let’s assume we have a sentence with a sentence length score(its one of our feature) of 0.3. To fuzzify this score, we calculate the degree of membership for each linguistic term using the triangular membership functions.
  For the LOW term: Degree of membership = 1 - (0.3 - 0)/(0.4 - 0) = 0.75.
  For the MEDIUM term: Degree of membership = (0.3 - 0.2)/(0.4 - 0.2) = 0.5.
  For the HIGH term: Degree of membership = (0.3 - 0.2)/(0.4 - 0.2) = 0.5.
  So, the fuzzified values for this sentence’s length would be:
  LOW: 0.75
  MEDIUM: 0.5
  HIGH: 0.5



Rule Base:

- Once the scores have been fuzzified into linguistic terms, fuzzy rules are applied to determine the importance of each sentence.

- Fuzzy rules are predefined statements that relate the input (linguistic terms) to the output (importance level) based on expert knowledge or predefined heuristics.

- The fuzzy rules can take various forms and depend on the specific context and requirements of the text summarization system. These rules can be defined using IF-THEN statements or other logical expressions.

- For instance, a fuzzy rule might state:

IF (NoWordInTitle is VH) and (SentenceLength is H) and (TermFreq is VH) and (SentencePosition is H) and (SentenceSimilarity is VH) and (NoProperNoun is H) and (NoThematicWord is VH) and (NumbericalData is H) THEN (Sentence is important)
By applying the fuzzy rules to the fuzzified scores, the system assigns an importance level (unimportant, average, or important) to each sentence

Fuzzy Inference:

a. Apply the fuzzy rules to determine the overall importance of each sentence based on the fuzzy inputs.

```
-----------------------------------------------------
Sentence:
This is a sample sentence.
Consequence: 0.8714964095499695

Sentence: Sentence similarity is calculated
between this sentence and the previous one.
Consequence: 0.7831926155125001

Sentence: Numerical data such as 123, 456, and 789
are extracted from this sentence.
Consequence: 0.7766478724551779

Sentence: The sample title is about natural language processing.
Consequence: 0.75

Sentence: Thematic words related to the topic are identified.
Consequence: 0.7491416702800158

Sentence: The sentence position score is calculated.
Consequence: 0.6

Sentence: The sentence length is calculated.
Consequence: 0.5

Sentence: The term weight of this sentence is calculated.
Consequence: 0.35
```

Text Summarization using Feature weights by Fuzzy:

- Feature weights play a pivotal role in determining sentence importance during our summarization process.

- These weights allow us to control the significance of individual features in the overall sentence ranking.

- We consider both Title Feature Scores and Thematic Word Feature Scores equally important, highlighting their alignment with the main title and the presence of thematic words as key factors in sentence scoring.

- Sentence similarity is given higher weight, as it reflects the sentence's alignment with the entire text, often indicative of vital information.

- Sentence length outweighs numerical Score Ratios as we aim for concise summaries, although both can carry equal weight when pinpointing crucial details.

- Utilizing the max membership method, we calculate each sentence's score by extracting the highest membership value for each feature, multiplying it by its respective weight, and summing these values.

- The top-ranked sentences, following this scoring, are selected for the final summary.

Optimization-of feature weights

•Optimization is done using the genetics algorithm.

•We specifically here utilize the real coded genetics algorithm not the binary coded because we have the real values so we don’t want to convert them again into binary and then again into decimal values.

•For the fitness function, we are using the F-Scores.

•Parent Selection: Selecting 2 individuals from the current population and making them parents for the next generation. Selecting a subset of superior individuals and taking the best of them to be the parent.

•Crossover:SBX crossover is used because the naïve crossoveroperators
such as the single point crossover might fail to perform well.

•We combined thegenerated offspring from the mutation and initialpopulation and then we arrange them according to their fitness values.

Comparison

•The main difference between the two methods lies in their application of fuzzy logic:
• In the first method, fuzzy rules guide the summarization decisions.
• In the second method, fuzzy logic is used to handle uncertainty in the feature values and determine the overall sentence scores.
•The second method is easier to implement in comparison to fuzzy rules but it involves more human involvement like in the calculation of feature weight which requires great domain knowledge. It also requires the user to decide which membership function is used, along with linguistic variables etc for each feature like the fuzzy rules.
•We have used first method because it is calculating the sentence score with less human involvement.

Evaluation Method : F-score

F-scores compare our model with the others. F-score, which is a combination of precision and recall, provides a balanced measure of the summarization system's performance.

1.True Positives (TP)

2.False Positives (FP)

3.False Negatives (FN)

4.Precision = TP / (TP + FP)

5.Recall = TP / (TP + FN)

6.F-score = 2 * (Precision * Recall) / (Precision + Recall)

- To calculate these metrics for text summarization, we need a reference summary (gold summary) and the generated summary produced by our summarization system.

Experiments

- The content of the generated summary is comared to the reference summary to count the metrics.
- So our dataset is basically the random texts of 20 lines paragraphs from the internet and their human-generated will act as gold summaries .
- Fuzzy inference model using Fuzzy rules is much better than the other two.
- Though in the fuzzy method using features weights, we have done optimization but due to early convergence we can't achieve maximum f-score.

f-score	*If-idf*	*Fuzzy (using feature weights)*	*Fuzzy(using fuzzy rules)*
1.	0.20	0.42	0.44
2.	0.54	0.56	0.66
3.	0.24	0.42	0.58

Improvements

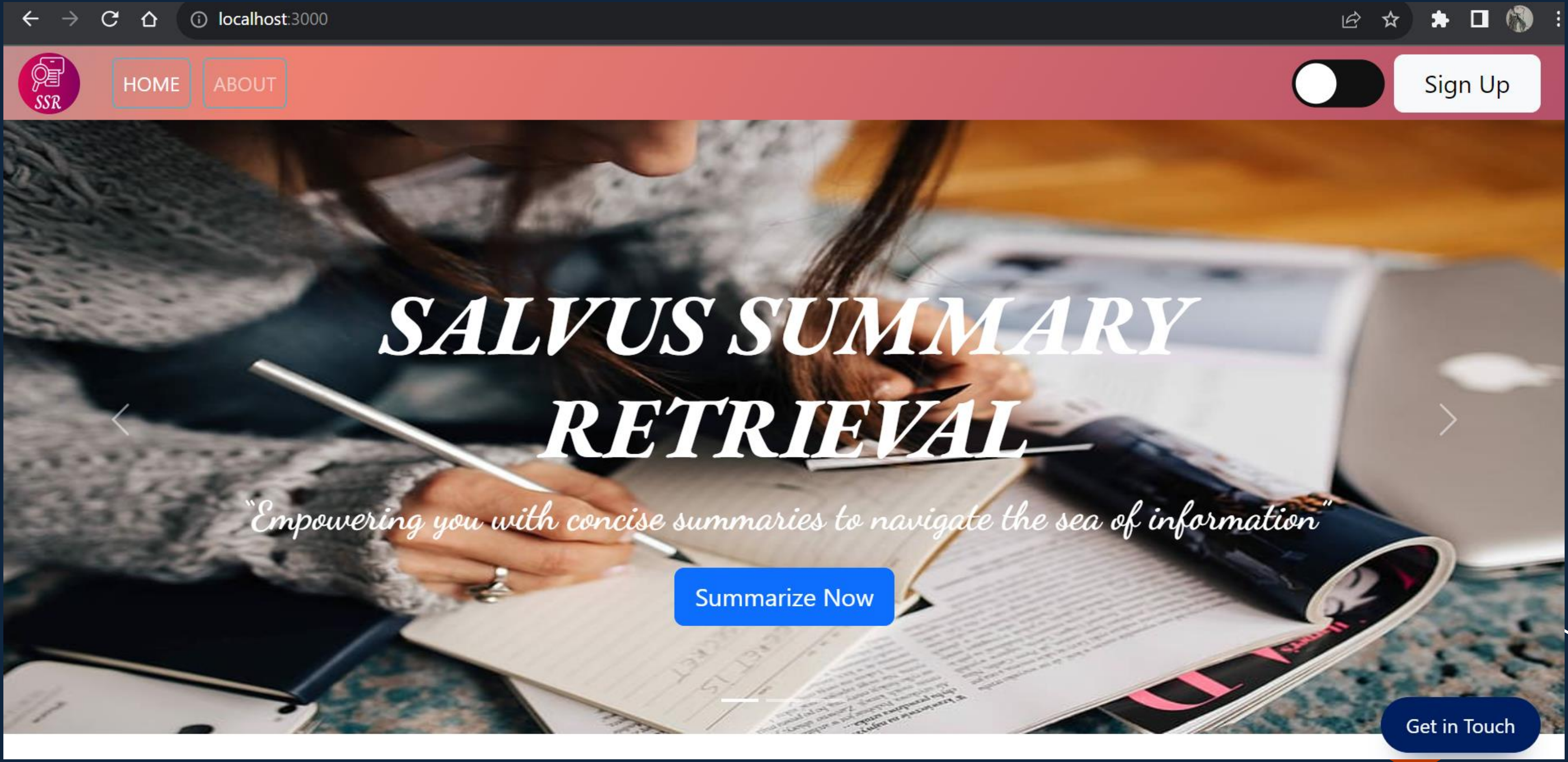
In the text summarization based on fuzzy rules, we can improve it further in the following ways-

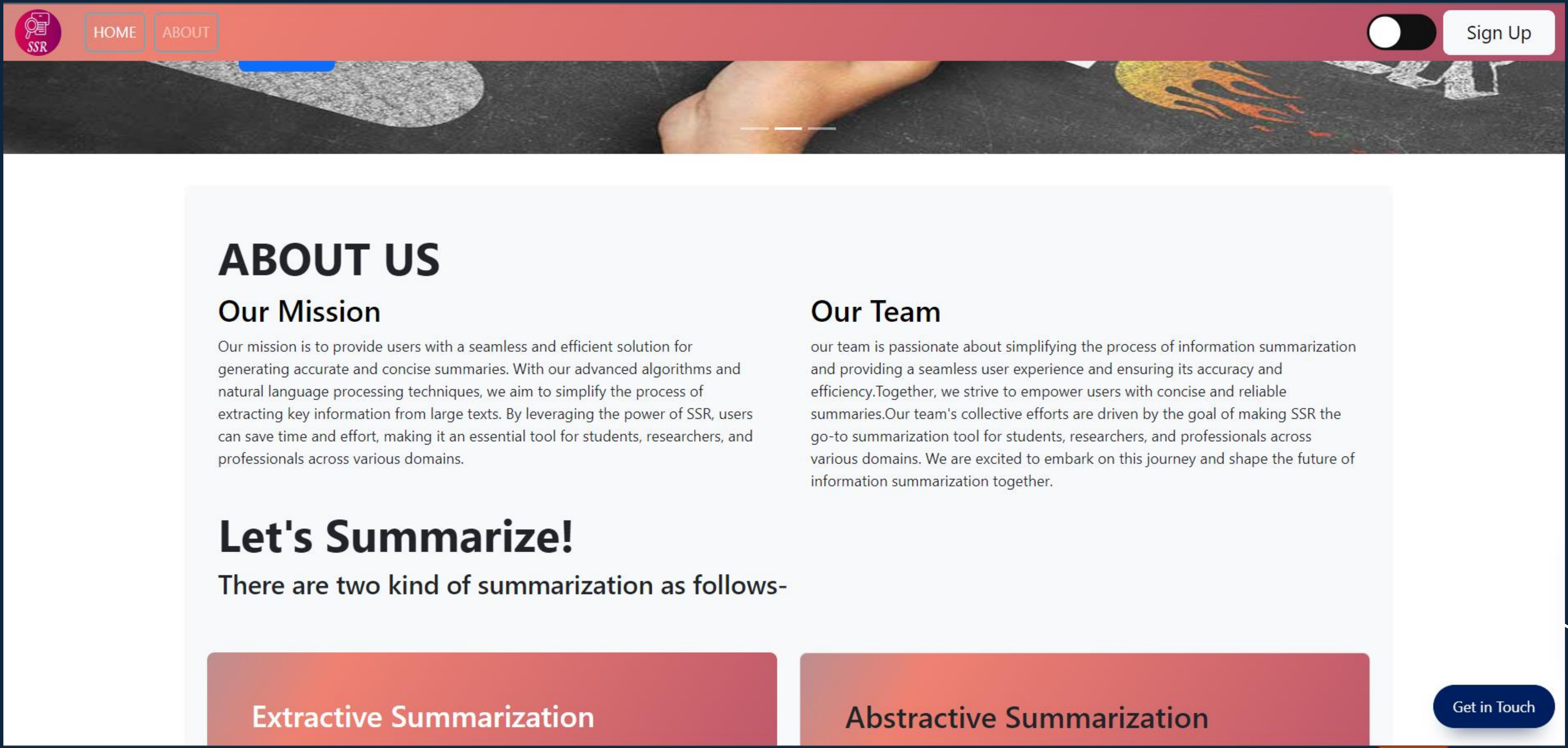
- Assigning the appropriate membership functions for each feature according to the domain.
- Defining different linguistic variable sets for different features. Also, while defining the boundaries of the membership functions.
- we can be more precise and further optimize them according to each feature.

In the text summarization-based feature weights, we can improve it further in the following ways-

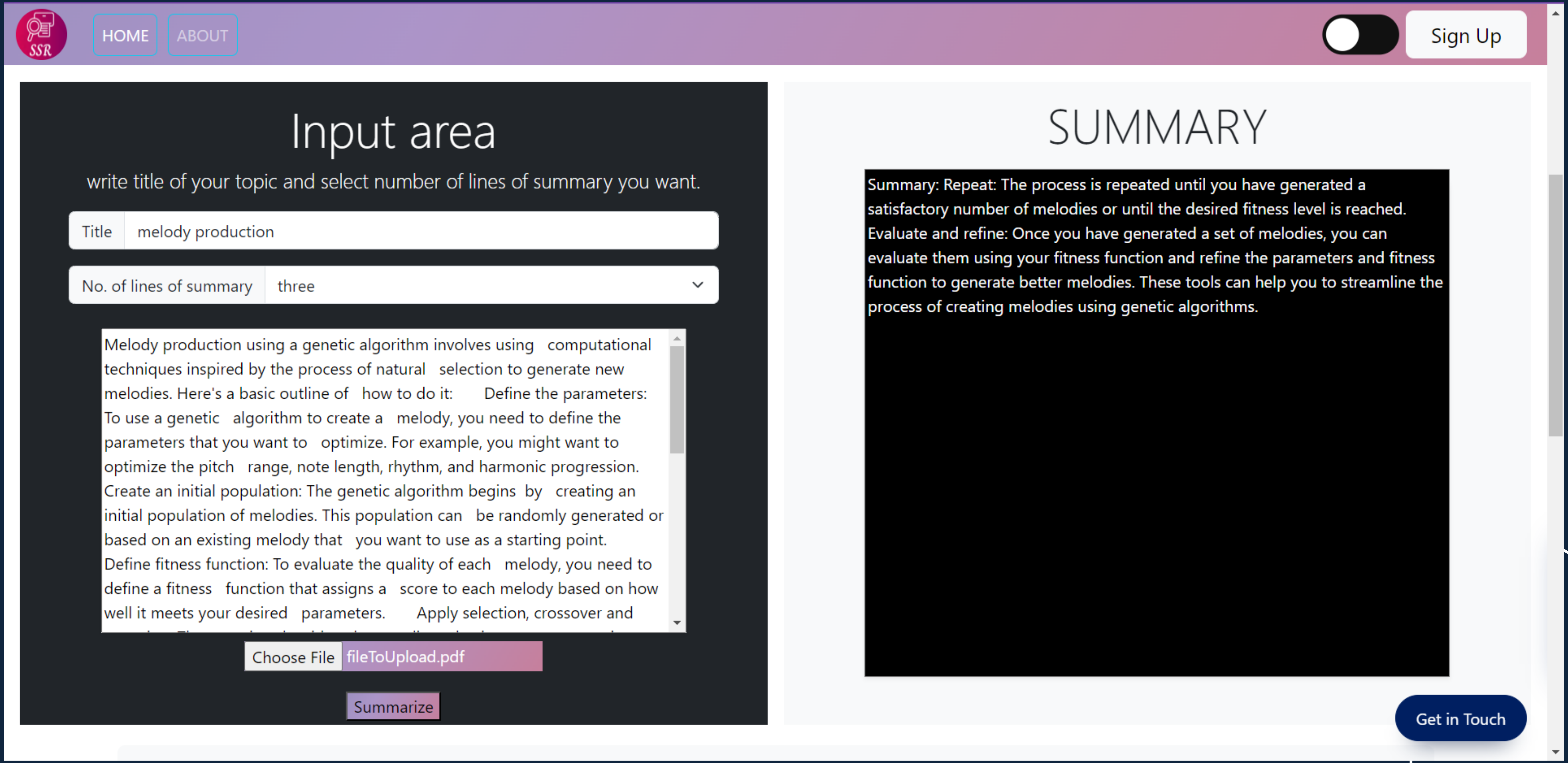
- We can optimize it further using the same methods as discussed above for text summarization based on fuzzy rules.
- Apart from that, most importantly, we can optimize the feature weights.
- By assigning different weights to various features, the model can be fine-tuned to produce summaries that align with the desired characteristics or priorities of the summarization process.

- Designed a rule-based system, defining fuzzy rules that relate input linguistic terms to output sentence importance levels.
- Employed fuzzy inference methods to aggregate fuzzy outputs, determining the importance of each sentence in the context of the entire document.
- Optimized feature weights using a genetic algorithm, fine-tuning the summarization process and achieving improved performance.
- Evaluated the performance of different text summarization models using the F-score measure, demonstrating that the fuzzy inference model with fuzzy rules outperformed other models.
- Explored avenues for further improvements, including the addition of new features, fine-tuning membership functions, and considering hybrid approaches to enhance summarization capabilities.
- These sentences provide a concise overview of the tasks completed in your research and highlight the key achievements in your project related to text summarization.


localhost:3000
SSR
HOME
ABOUT
Sign Up
SALVUS SUMMARY RETRIEVAL
"Empowering you with concise summaries to navigate the sea of information"
Summarize Now
Get in Touch


SSR
HOME
ABOUT
Sign Up
ABOUT US
Our Mission
Our mission is to provide users with a seamless and efficient solution for generating accurate and concise summaries. With our advanced algorithms and natural language processing techniques, we aim to simplify the process of extracting key information from large texts. By leveraging the power of SSR, users can save time and effort, making it an essential tool for students, researchers, and professionals across various domains.
Our Team
our team is passionate about simplifying the process of information summarization and providing a seamless user experience and ensuring its accuracy and efficiency.Together, we strive to empower users with concise and reliable summaries.Our team's collective efforts are driven by the goal of making SSR the go-to summarization tool for students, researchers, and professionals across various domains. We are excited to embark on this journey and shape the future of information summarization together.
Let's Summarize!
There are two kind of summarization as follows-
Extractive Summarization
Abstractive Summarization
Get in Touch

Extractive Summary



Conclusion

The results show that fuzzy logic may be successfully applied to text summarization.

We used fuzzy methods to precisely determine sentence relevance through data pre-processing and feature extraction.

Our optimisation efforts, which included feature weight modification using genetic algorithms, significantly improved the results of the summarization.

We overcame the uncertainty that comes with natural language by using fuzzy logic, opening the way for more accurate and detailed text summary.

These results highlight how fuzzy-based methods may improve a range of applications, including knowledge extraction and information retrieval.

References

[1] Fattah, M. A. and Ren, F.: 2008, Automatic text summarization, Proceedings of World Academy of Science, Engineering and Technology, Vol. 27, pp. 192–195.

[2] Lin, C.-Y.: 2004, ROUGE: A package for automatic evaluation of summaries, Proceedings of Workshop on Text Summarization of ACL, Spain.

[3] Luhn, H. P.: 1958, The automatic creation of literature abstracts, IBM Journal of Research and Development 2, 159–165

[4] Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, Hoda K. Mohamed, Automatic text summarization: A comprehensive survey, Expert Systems with Applications, Volume 165, 2021,113679,ISSN 0957-4174.

Thank you