GENETIC ALGORITHM FOR TEXT SUMMARIZATION

Anushka Sagar

Anushkasagar302@gmail.com

Department of Computer Science and Engineering, Lovely Professional University, Punjab, India

ABSTRACT

This abstract presents a text compression process for producing shorter document. To quickly access the important goals and main features of the input document. This project uses the application of genetic algorithms to summarize a text to automatically generate shorter and informative summaries from large volume of text. This approach represents text data as chromosomes and uses genetic operators such as selection, crossover and mutation to evolve population of summaries. The fitness of each summary is evaluated based on criteria including information, coherence and length, results in producing high qualities summaries.

1 Introduction

Due to huge information, it is very necessary to efficiently to reduce vast amounts of text into concise and meaningful summaries. Text summarization techniques aim to address this challenge by automatically generate the concise versions of textual content while preserving its essential information. Genetic algorithms work by drawing inspiration from principles of natural selection and evolution.

Genetic algorithm are optimization techniques that emulate the process of natural selection to iteratively evolve solution to complex problem. By representing potential selection to iteratively evolve solutions to complex problems. By representing potential selection as chromosomes and applying genetic operators such as selection, crossover and mutation.

Overall, this study seeks to contribute to ongoing efforts in advancing text summarization techniques by exploring the innovative application of genetic algorithms.

2 Methodology

2.1 Chromosome Representation

Represent each potential summary as a chromosome like binary strings, array or other data structure where each gene corresponds to a sentence or a word.

2.2 Population Initialization

Generate an initial population of summaries randomly or using heuristic method.

2.3 Fitness Function

Define a fitness function to evaluate how good a summary is. This function account metrics like informativeness, coherence, and length of the summary.

2.3 Selection

this process used to select or choose the individual from the population based on their fitness function.

2.4 Crossover

Apply crossover operations to select individuals to produce offspring. For text summarization, we can perfume crossover at sentence or word level to combine information from different summaries.

2.5 Mutation & Replacement

Name: Anushka Sagar

Code: file&code

Mutation: Introduce variation into the population by applying mutation operations. This could involve randomly changing words or sentences in a summary.

Replacement: Replace individuals in the current population with the offspring generated through crossover and mutation.

2.6 Termination & Result

Termination Condition: Decide when to stop the algorithm. This could be after a certain number of generations or when the best summary meets certain criteria.

Result Extraction: Extract the best summary from the final population based on the fitness function.

2.7 Evaluation

Evaluate the quality of the final summary using human judgment or automatic evaluation metrics.

3 Results & discussion

3.1 Dataset

The dataset used in this Genetic algorithm for text summarization project. The dataset consist of Total `length of the text in /content/sumtext.txt: 633.`

3.2 Qualitative results

The performance of Genetic algorithms like Preprocessing, Understanding the Text, Feature Extraction, Summarization Techniques, Evaluation. After using these techniques on our text data, we got a concise and informative text summary of `Length of text after summarization: 28.`

4 Conclusion

In summary, text summarization using genetic algorithms offers an automated approach to distil large amounts of text into concise summaries. Leveraging principles from natural selection, these algorithms iteratively refine summaries by representing them as chromosomes and applying genetic operations like selection, crossover, and mutation.

While promising, challenges include parameter tuning, potential convergence to suboptimal solutions, and scalability issues with large datasets. Nonetheless, ongoing research and advancements in genetic algorithms, combined with other techniques, continue to enhance text summarization systems' effectiveness across various domains.

5 References

1. **Paper: "A Novel Approach to Automatic Text Summarization Using Genetic Algorithms" by Mohammed Al-Ma'adeed, Volodymyr Ponomaryov, and Vladimir Dobriy.**
2. **Paper: "Text Summarization Using Genetic Algorithms" by Saqib Mahmood and Qasim Mahmood.**
3. **Paper: "Genetic Algorithm based Text Summarization: A Review" by Pradeep Kumar Singh and P. K. Mishra.**