Scientific Document Summarization using Citation Context and Multi-objective Optimization

Naveen Saini*, Sushil Kumar, Sriparna Saha, And Pushpak Bhattacharya
Indian Institute Of Technology Patna, Bihar, India-801106

ABSTRACT

Instead of using all sentences present in the article, we have considered only those sentences which are relevant to the citation contexts. Therefore, as the first step, the relevant sentences in the reference paper (RP) are identified in the semantic space and then, the following algorithm is applied.

ALGORITHM

BACKGROUND

Multi-objective optimization
- Optimization of more than one objectives simultaneously
- Provide a set of Pareto optimal solutions in a single run
- Users can choose any one based on his/her choice

Evolutionary algorithms
- Meta-heuristic optimization algorithms to find an optimal solution
- Inspired by biological phenomenon (crossover, mutation and selection) in the natures
- Differential evolution (DE) is one such algorithm.
- It starts from a set of solutions (called as population)
- Each solution is associated with some (subjective) values to measure their quality
- These solutions evolves over the iteration to generate new population
- Only those solutions proceed to the next generation which are good in terms of “survival of the fittest” principle.

OBJECTIVE FUNCTIONS

- F1: Sentence position in the article (F1↑) (↑)
- F2: Maximum similarity with the Title (F2)
- Reciprocal of Word mover distance utilizing GoogleNews word2vec model (↑)
- F3: Maximum overlap with the top-scoring sentences provided by LexRank algorithm (F3↑) (↑)

Thus, objective is to maximize these objective functions represented as

max { F1↑, F2, F3↑ }