Scientific Document Summarization using Citation Context and Multi-objective Optimization

(Paper ID: 1202)

Authored by:

Naveen **Saini**, Sushil **Kumar**, Sriparna **Saha**, And Pushpak **Bhattacharyya**

Indian Institute Of Technology Patna, Bihar, India-801103

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

Milan, Italy 10 | 15 January 2021

Presenter: Naveen Saini

Outline

- Definition of Summarization
- Why Summarization??
- Approaches of Summarization
- Example of Summarization
- Application of Summarization
- Scientific Document Summarization
- Preliminaries
  - Multi-objective optimization
  - Evolutionary algorithm
- Proposed Methodology
- Results and discussion

Summarization

- Task of automatically creating a compressed version of the text document (set of tweets, web-page, single/multi-document) that should be relevant, non-redundant and representative of the main idea of the text.

- A text that is produced from one or more texts that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that.

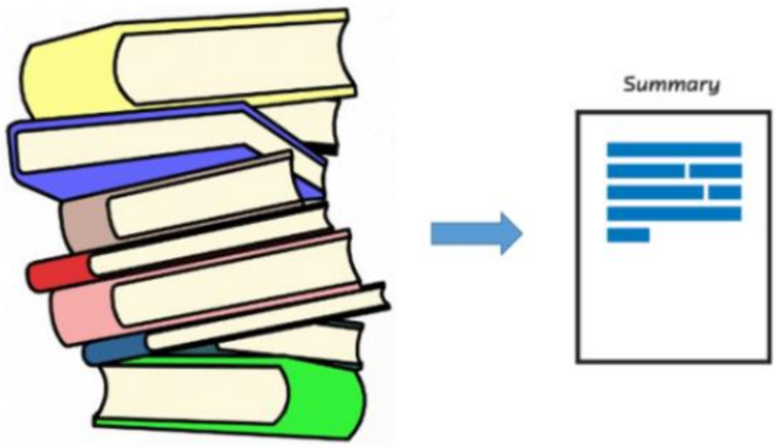


Image credit: Wikipedia

Why summarization?

- Internet has provided large collections of text on a variety of topics
- large number of electronic documents are available online

Problems

- Users get so exhausted reading large amount
- User face difficulty in finding relevant information

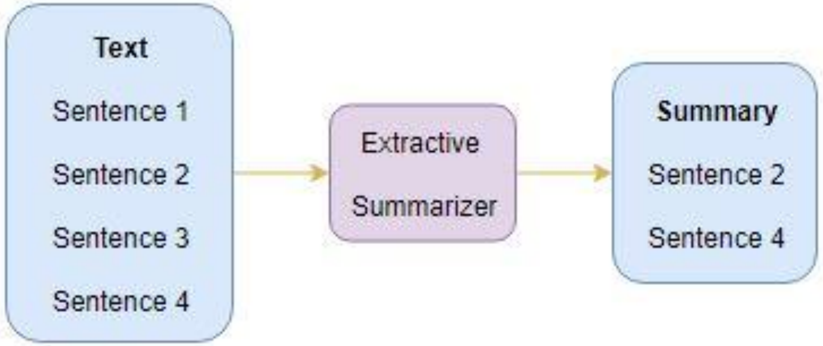
Solution

- Automatic text summarization system is needed that compress information into shorter length that must follow coverage of information, non-redundancy, information significance and Cohesion in the text

Approaches of Summarization

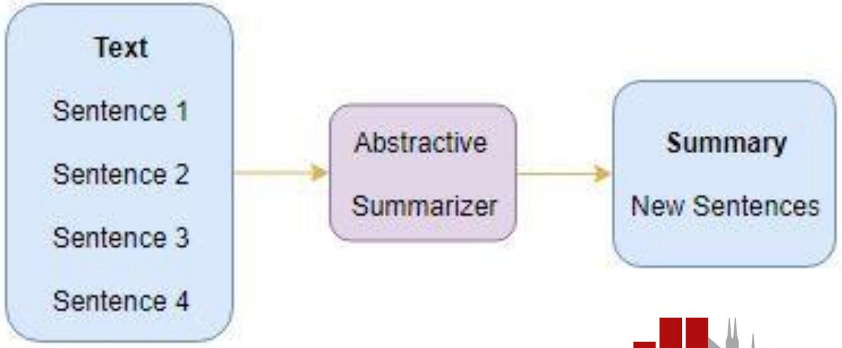
Extractive

- Selecting a few relevant sentences from the original document
- Relevance of sentences is decided using sentence scoring features like sentence position, similarity with the title etc.



Abstractive

- Abstract summary which includes words and phrases different from the ones occurring in the source document
- Required natural language generation or reconstruction of sentences



Examples of Summarization

[English is the dominant language in the writing and publishing of scientific research in the form of scientific articles.]$_1$ [However, many non-natives users of English suffer the interference of their mother tongues when writing scientific papers in English.]$_2$ [These users face problems concerning rules of grammar and style, and/or feel unable to generate standard expressions and clauses, and the longer linguistic compositions which are conventional in this genre.]$_3$ [In order to ease these users' problems, we developed a learning environment for scientific writing named AMADEUS (Amiable Article Development for User Support).]$_4$ [AMADEUS consists of several interrelated tools reference, support, critic and tutoring tools and provides the context in which this dissertation is inserted.]5 [The main goal of this research is to implement AMADEUS as an agent -based architecture with collaborative agents communicating with a special agent embodying a dynamic user model.]$_6$ [In order to do that we introduce the concept of adaptivity in computer systems and describe several user model shells.] $_7$ [We also provide details about intelligent agents which were used to implement the user model for the AMADEUS environment.]$_8$

extractive

English is the dominant language in the writing and publishing of scientific research in the form of scientific articles. In order to ease these users' problems, we developed a learning environment for scientific writing named AMADEUS (Amiable Article Development for User Support). The main goal of this research is to implement AMADEUS as an agent - based architecture with collaborative agents communicating with a special agent embodying a dynamic user model. We also provide details about intelligent agents which were used to implement the user model for the AMADEUS environment.

A detained iranian-american academic accused of acting against national security has been released from a tehran prison after a hefty bail was posted, a to p judiciary official said tuesday.

abstractive

iranian-american academic held in tehran released on bail.

Application of Summarization

S.No.	Types of summary	Factors
1	query-focused	Summarize a document given a query
2	Single/Multi-document	For summarizing single or multiple documents
3	Figurative	For summarizing figures
4	Web-based	For summarizing web pages
5	E-mail based	For summarizing e-mails
6	Personalized	Information specific to a user's need
7	Sentiment-based	Opinions are detected
8	Microblogs	For summarizing a set of tweets
9	Scientific Document Summ.	For summarizing scientific documents

NOTE:

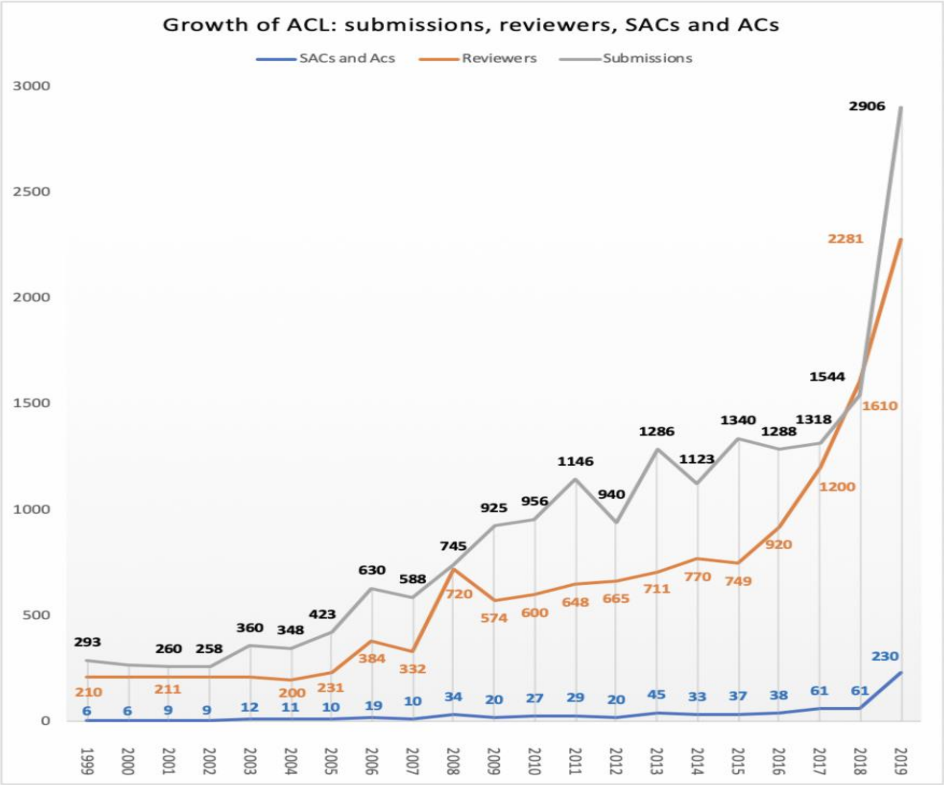
- Any system can be developed either in supervised or unsupervised way.

Scientific Document Summarization

- Rapid growth in the field of scientific literature → challenging task for the researchers to gain <u>up-to date knowledge of the current advancements</u>.

- In Figure, growth of ACL (a top tier NLP conference) is shown

- Solution: Summarizing the scientific article address this challenge.



SAC: Senior Area Chair; AC: Area Chair

Presenter: Naveen Saini

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Queries ??

Abstract???

Can we consider abstract as a summary?

(a) covers only a broad idea of the article
(b) may lack of detailed contributions

Abstract...

Intro ....

Method......

Results.........

Conclusion...

Reference paper

Citation Context???

Can we consider citation contexts as a summary???

(a) written by different authors
(b) may be inaccurate or misunderstood by the authors

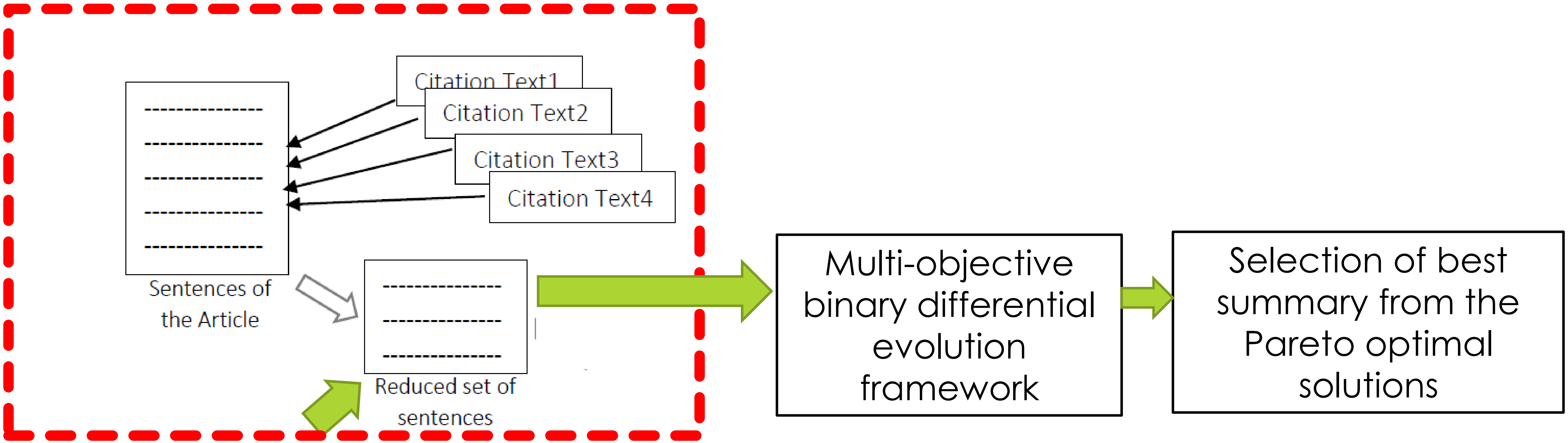
How to Address these challenge??

INDIAN INSTITUTE OF TECHNOLOGY PATNA

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Scientific Document Summarization via Citation Context

- The possible way to address these challenges is to summarize the RP using the citation contextualization



MULTI-OBJECTIVE OPTIMIZATION (MOO)

Optimization of two or more than two objectives simultaneously

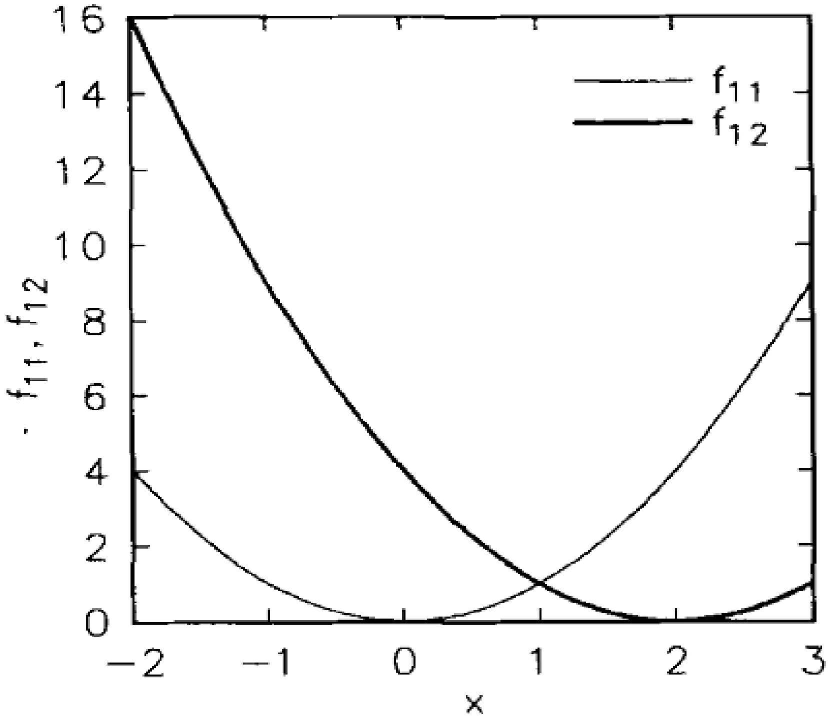
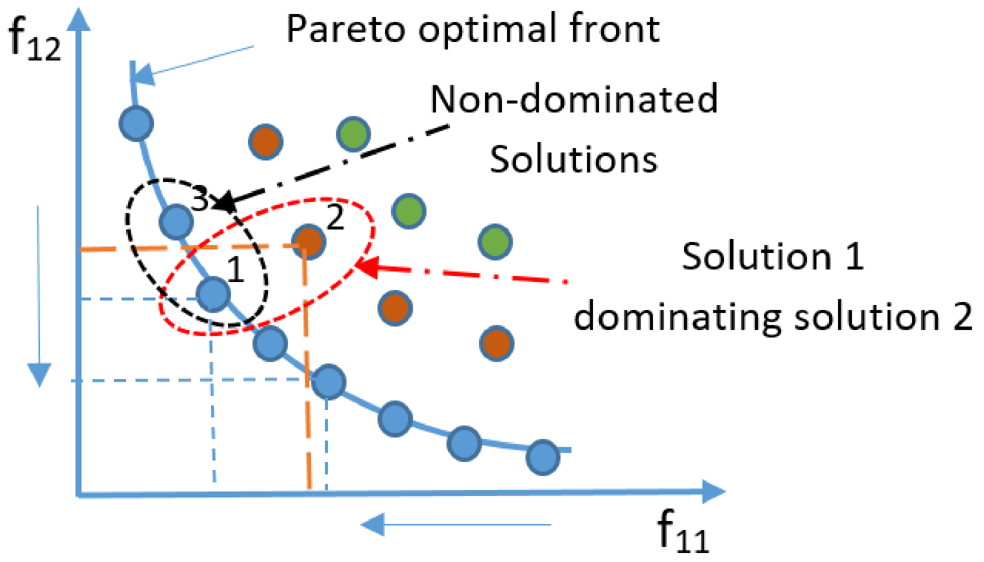
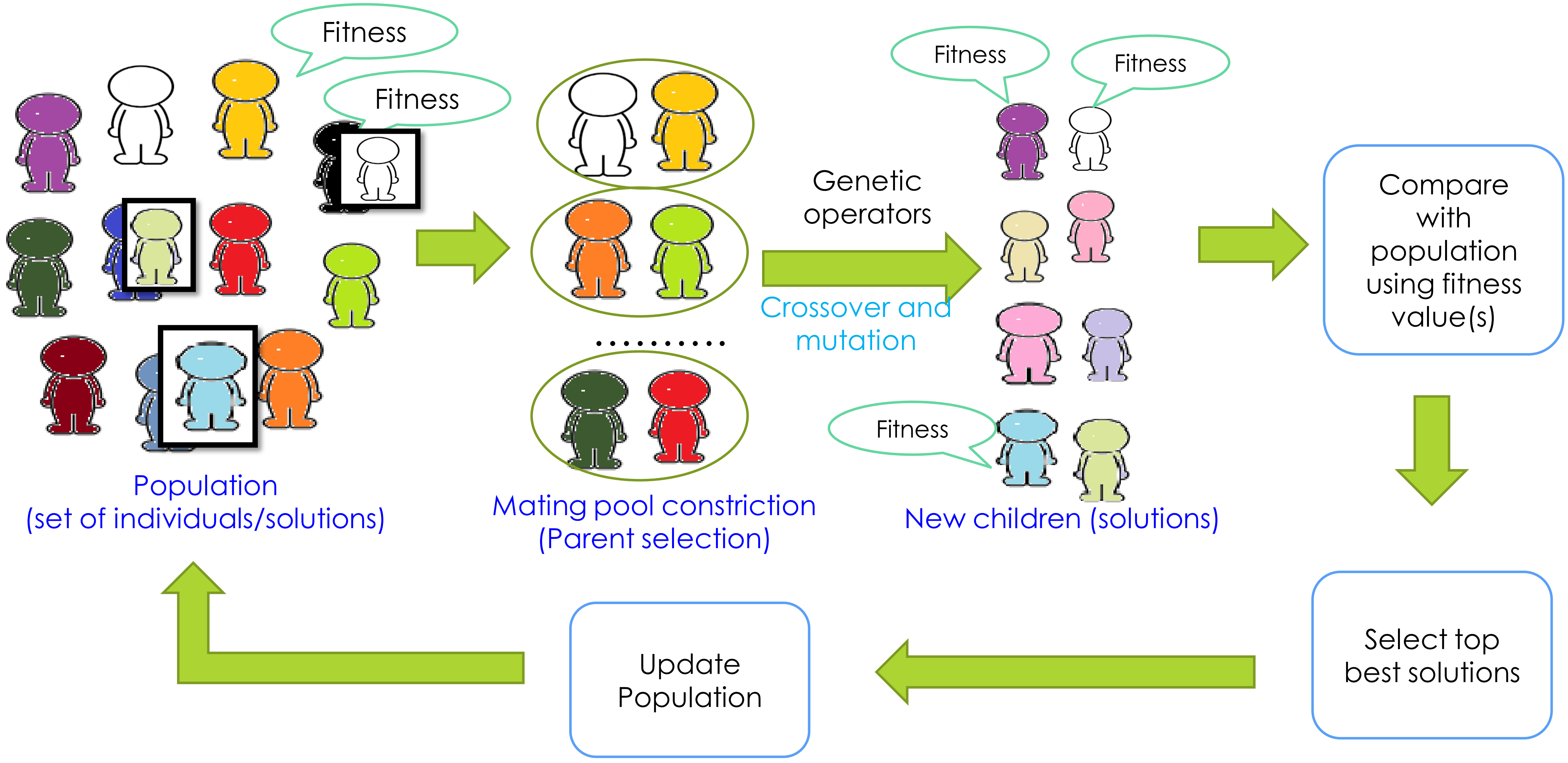


Image credit: Wikipedia

- The solution x = 0 is optimum w.r.t. function but not so good with respect to f_{12}.
- The solution x = 2 is optimum w.r.t. function f_{12} but not with respect to function f_{11}.
- <u>Optimal range of values</u> of both functions: $0 <= x <= 2$ which provides a set of solutions.

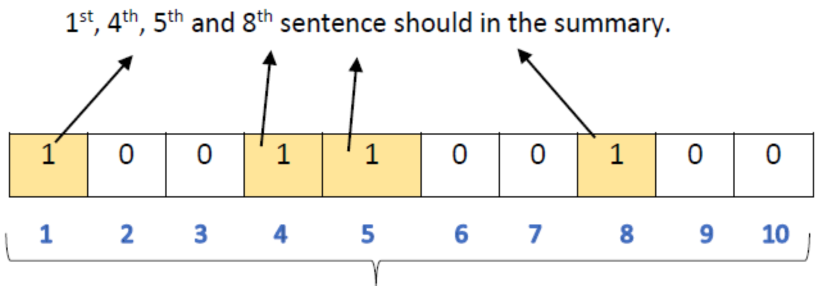


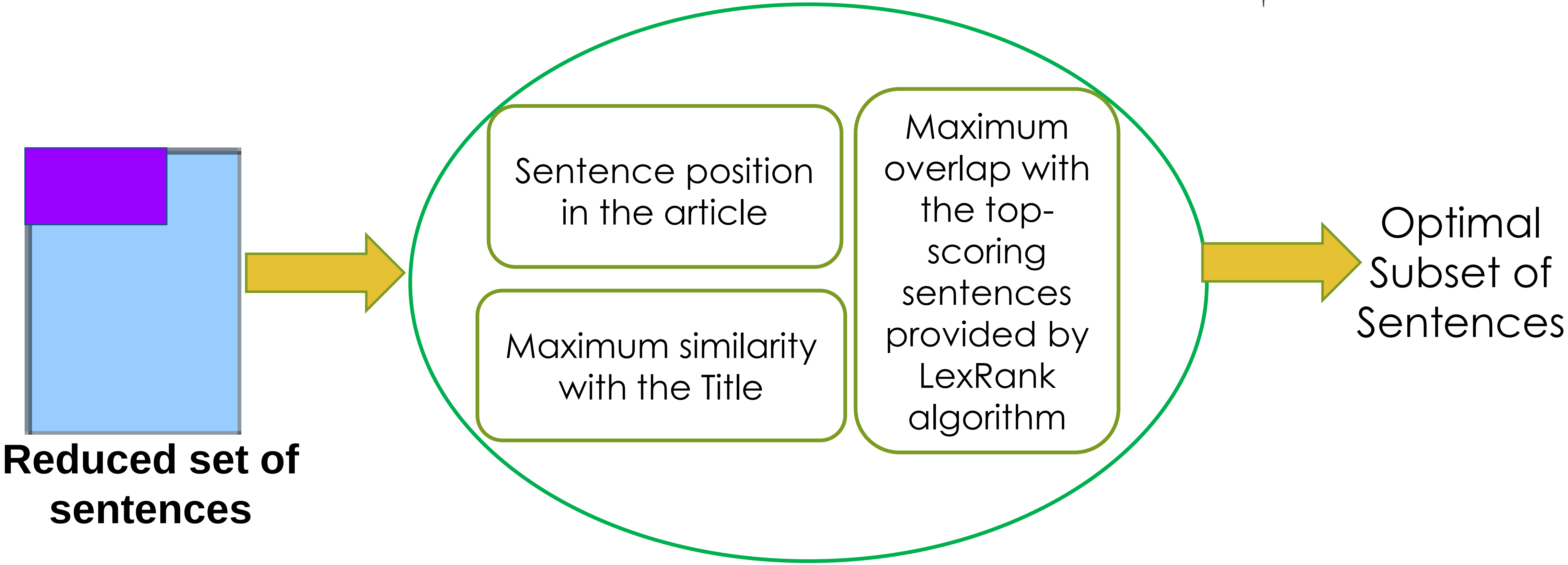
EVOLUTIONARY ALGORITHM (EA)



Crossover: Exchange of genes; Mutation: change in the gene value

Problem Formulation





Sentence scoring functions: Optimize using multi-objective binary differential evolution

Objective functions

Objective functions	Description
Sentence position in the article (F1) (↑)	It has been shown that starting sentences of the articles or leading sentences of the paragraphs convey relevant information. Therefore, this feature has been explored for the SDS task.

Maximum similarity with the Title (F2)
 	
For (a): (↑)
For (b): (↓)
	It is quite obvious that the title of any article available in the form of blogs, news, documents, among others, describes the theme of the article. Therefore, the feature 'sentence similarity with the title' is explored for SDS task. Two ways of measuring similarity, each differs in utilizing word representation and similarity measure: (a) fast-text, cosine similarity; (b) Googlenews word2vec, word mover distance.

Maximum overlap with the top-scoring sentences provided by LexRank algorithm (F3) (↑)	It counts the number of overlapping sentences with the top-scoring sentences identified by LexRank algorithm.

Proposed Methodology

Algorithm 1 Procedure of MOOTweetSumm+

1: $\mathbb{P} \leftarrow$ Initialize Population $< X^1, X^2, X^3, \ldots, X^{|\mathbb{P}|} >$
2: For each solution $\mathcal{X}$, evaluate objective functional values
3: $CGen$=0 ▷ Current generation number
4: Repeat step-5 to 9 until $CGen < MaxGen$
5: $\mathbb{P}^{'}$=[] ▷ Population to store new solutions
6: For each solution $\mathcal{X} \in \mathbb{P}$, generate new solution

(a) Randomly select three solutions $r1, r2$ and $r3$ from $\mathbb{P}$ to form a mating pool
(b) $Prob(\mathcal{X}) \leftarrow$ Perform probability estimation operator using selected random solutions and $\mathcal{X}$
(c) $Y^{'} \leftarrow$ Convert $Prob(\mathcal{X})$ into a binary solution
(d) $Y^{''} \leftarrow$ Perform crossover between $Y^{'}$ and $\mathcal{X}$
(e) Evaluate objective functions for $Y^{''}$
(f) Add $Y^{''}$ into $\mathbb{P}^{'}$

7: Merge Old population ($\mathbb{P}$) and new population ($\mathbb{P}^{'}$)
8: $\mathbb{P} \leftarrow$ Select the best $|\ \mathbb{P}\ |$ solutions based on their objective functional values using non-dominating sorting and crowding distance operator
$CGen \leftarrow CGen$+1
return the best summary

$$P(x_j^t) = \frac{1}{1 + e^{-\frac{2b \times \left[x^t_{r1,j} + F \times \left(x^t_{r2,j} - x^t_{r3,j}\right) - 0.5\right]}{1+2F}}}$$

$$y_j' = \begin{cases} 1, & \text{if rand()} \leq P(x_j^t) \\ 0, & \text{otherwise} \end{cases}$$

$$y_j'' = \begin{cases} y_j', & \text{if rand()} \leq \text{CR} \\ x_j, & \text{Otherwise} \end{cases}$$

Key-contributions

- First attempt to develop an unsupervised extractive summarization technique for solving scientific document summarization task using an evolutionary algorithm (EA).

- Existing works have considered only a single trait/objective function for generating a summary, but here, multiple traits are utilized in generating summary, and those traits are simultaneously optimized using multi-objective optimization (MOO) concept.

- In existing works on SDS, the graph-based feature has never been explored for calculating the sentence relevance scores. Therefore, a graph-based feature utilizing the LexRank algorithm has been explored in integration with our MOO-based framework

- For a given dataset, the selection of appropriate objective functions which can be optimized simultaneously in obtaining a good quality summary is a crucial task; therefore, an ablation study has been performed on the usage of various objective functions to identify the best candidate set of sentence-scoring functions

Evaluation measure

CL-SciSumm→	**2016**	**2017**
Test	10	10
Train	10	30
Dev	10	0
Avg. #Citations	35.4	16.1

Datasets: CL-SciSumm 2016 and CL-SciSumm 2017

ROUGE-N:

$$ROUGE-N = \frac{\sum_{S \in Summary_{ref}} \sum_{N-gram \in S} Count_{match}(N-gram)}{\sum_{S \in Summary_{ref}} \sum_{N-gram \in S} Count(N-gram)}$$

Where N represents the length of n-gram, Countmatch(N −gram) is the maximum number of overlapping N −grams between reference/gold summary and system summary, Count(N −gram) is the total number of N − gram in the reference summary. In our experiment, N takes the values 2 for ROUGE−2.

Results (1/2)

TABLE II
COMPARISON OF OUR PROPOSED APPROACHES WITH TOP SUBMITTED SYSTEMS OF CL-SCISUMM 2016 SHARED TASK.

Type of Summary→	Human	Community	Abstract
Method	**ROUGE-SU4**		
Ours ($F1$, $F2_w$)	**0.24**	**0.41**	**0.45**
Ours ($F1$, $F2_v$)	**0.25**	**0.39**	**0.43**
Ours ($F2_w$, $F3$)	**0.21**	**0.36**	**0.43**
Ours ($F2_v$, $F3$)	**0.20**	**0.33**	**0.37**
sys8$PARA_7	0.14	0.13	0.42
sys8$PARA_1	0.11	0.13	0.25
sys3$LMKL1_CCS1	0.12	0.09	0.18
sys3$LMEQUAL_CCS2	0.12	0.10	0.21

TABLE III
COMPARISON OF OUR PROPOSED APPROACHES WITH TOP SUBMITTED SYSTEMS OF CL-SCISUMM 2017 SHARED TASK.

Type of Summary→	Human	Community	Abstract
Method	**ROUGE-SU4**		
Ours ($F1$, $F2_w$)	**0.26**	**0.38**	**0.27**
Ours ($F1$, $F2_v$)	**0.24**	**0.37**	**0.27**
Ours ($F2_w$, $F3$)	**0.25**	**0.35**	**0.22**
Ours ($F2_v$, $F3$)	**0.25**	**0.35**	**0.21**
CIST Run 3 [9]	0.17	0.16	0.17
UniMA Run 4,5,6 [20]	0.16	0.17	0.19
UniMA Run 7,8,9 [20]	0.16	0.16	0.18
Mean Score (all systems) [5]	0.14	0.14	0.15

- For CL-SciSumm 2016 dataset, our system attains improvements by 0.20, 0.28, and, 0.02 points over human, community and abstract summary of the best submitted system
- For CL-SciSumm 2017 dataset, our best system improves by 0.12, 0.24, and 0.12 points over the mean scores of all the submitted systems in the shared task.

Results (2/2)

TABLE IV
COMPARISONS BETWEEN BEST ROUGE SCORES ATTAINED BY PROPOSED METHODS WITH THE EXISTING METHODS ON CL-SCISUMM 2016 DATASET CONSIDERING HUMAN SUMMARY.

Human Summary		
Evaluation Measure→	**ROUGE-2**	**ROUGE-SU4**
Method↓	**F1-scores**	
Ours $(F1, F2_w)$	**0.23**	**0.24**
Ours $(F1, F2_v)$	**0.23**	**0.25**
Ours $(F2_w, F3)$	0.19	**0.21**
Ours $(F2_v, F3)$	0.19	**0.29**
LexRank[7]	0.12	0.11
CLexRank [8]	0.06	0.09
SumBasic [10]	0.09	0.12
CIST [9]	0.22	0.14
LMKL [11]	0.19	0.11
LMeq [11]	0.19	0.12
SUMMA [34]	0.13	0.09

TABLE V
COMPARISONS BETWEEN BEST ROUGE-2 SCORES (PRECISION (P), RECA (R) AND F1-SCORE (F1)) ATTAINED BY THE PROPOSED METHOD WITH T EXISTING METHODS ON CL-SCISUMM 2017 DATASET CONSIDERING ABSTRACT SUMMARY.

Abstract Summary			
Evaluation Measure→	**ROUGE-2**		
Method↓	**Avg. P**	**Avg. R**	**Avg. F1**
Ours $(F1, F2_w)$	**0.26**	**0.41**	**0.26**
Ours $(F1, F2_v)$	**0.26**	**0.40**	**0.26**
Ours $(F2_w, F3)$	0.20	**0.33**	0.20
Ours $(F2_v, F3)$	0.19	**0.34**	0.19
Word2vec [19]	0.21	0.27	0.24
tfidf-1:1 [19]	0.21	0.24	0.22
tfidf-1:3 [19]	0.20	0.24	0.21
tfidf-1:2 [19]	0.19	0.24	0.21
Lambdamart [33]	0.19	0.23	0.21
Filter [19]	0.18	0.23	0.20
Lambda [33]	0.26	0.02	0.20

Pareto Fronts

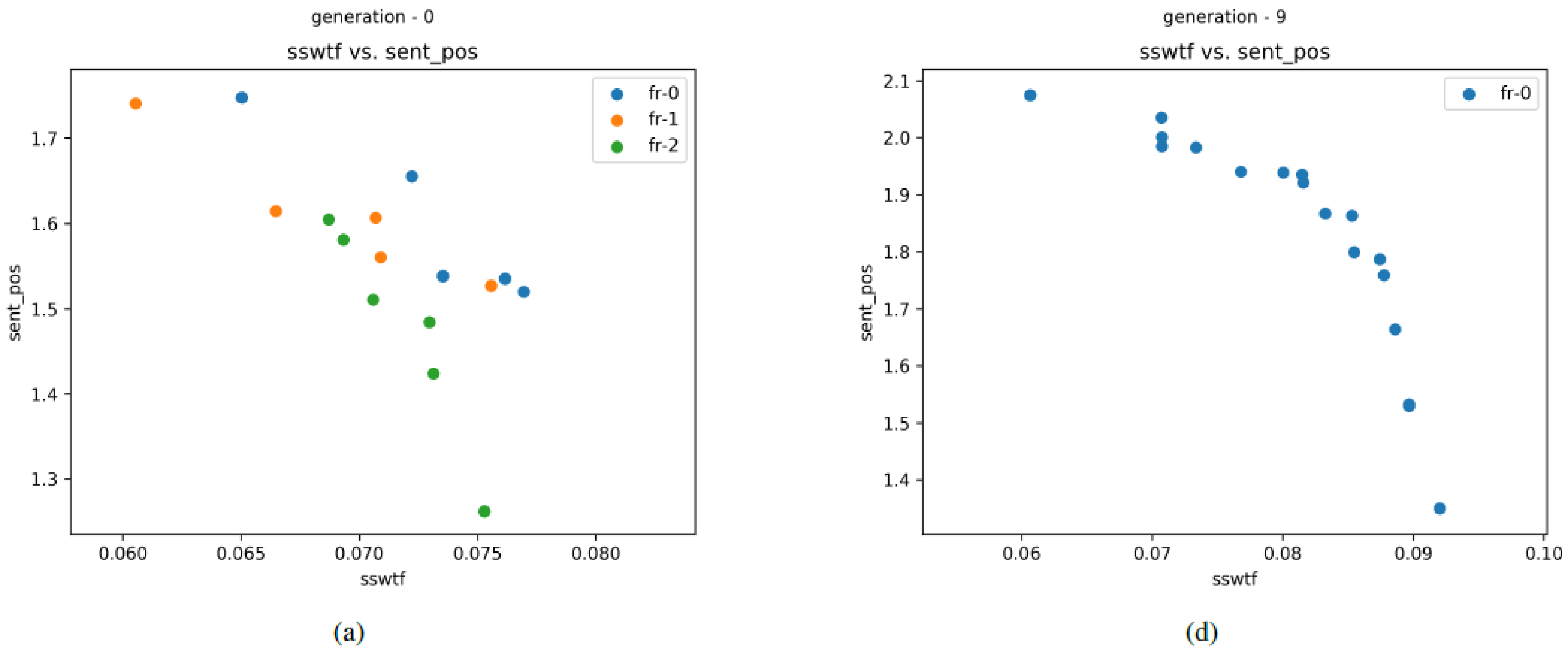


Fig. 3. Pareto Fronts generated by our proposed *MOOSciSumm* approach at the end of $0th$ and $9th$ generations after optimizing two objective functions (i) $SentPos$; (ii) $MaxSimWithTitle$ utilizing FastText vector. Here, sent_pos and $sswtf$ refer to the mentioned objective function in (a) and (b), respectively. Here, fr-0 in the legend indicates the solutions of rank-1 and so on.

TABLE VIII

THIS TABLE CONTAINS ONE OF THE BEST SYSTEM SUMMARIES GENERATED BY OUR PROPOSED APPROACH WITH ROUGE-2, ROUGE-SU4 SCORES OF 0.58 AND 0.60, RESPECTIVELY

Article No.: W11-0815 Predicted Community Summary:

The extensive use of Multiword Expressions (MWE) in natural language texts prompts more detailed studies that aim for a more adequate treatment of these expressions. This work consists in investigating the impact of Multiword Expressions on applications, focusing on compound nouns in Information Retrieval systems, and whether a more adequate treatment for these expressions can bring possible improvements in the indexing these expressions. One of the great challenges of NLP is the identification of such expressions, “hidden” in texts of various genres. For example, if the query was pop star meaning celebrity, and the terms were indexed individually, the relevant documents may not be retrieved and the system would 101 Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011), pages 101–109, Portland, Oregon, USA, 23 June 2011. One of the motivations of this work is to investigate if the identification and appropriate treatment of Multiword Expressions (MWEs) in an application contributes to improve results and ultimately lead to more precise man-machine interaction. The goal of our evaluation is to detect differences between the quality of the standard IR system, without any treatment for MWEs, and the same system improved with the identification of MWEs in the queries and in the documents. In this paper, we perform an application-oriented evaluation of the inclusion of MWE treatment into an Information Retrieval (IR) system. Along with the LEMA field, extracted in the previous procedure, we also extracted the value of the field POS (part-of-speech).

Community Gold Summary:

The extensive use of Multiword Expressions (MWE) in natural language texts prompts more detailed studies that aim for a more adequate treatment of these expressions. The automatic discovery of specific types of MWEs has attracted the attention of many researchers in NLP over the past years. One of the motivations of this work is to investigate if the identification and appropriate treatment of Multiword Expressions (MWEs) in an application contributes to improve results and ultimately lead to more precise man-machine interaction. This task aimed to explore the contribution of the disambiguation of words to bilingual or monolingual IR. In this paper, we perform an application-oriented evaluation of the inclusion of MWE treatment into an Information Retrieval (IR) system. Although language processing is not vital to modern IR systems, it may be convenient (Sparck Jones, 1997) and in this scenario, NLP techniques may contribute in the selection of MWEs for indexing as single units in the IR system. The selection of appropriate indexing terms is a key factor for the quality of IR systems. For example, if the query was pop star meaning celebrity, and the terms were indexed individually, the relevant documents may not be retrieved and the system would 101 Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011), pages 101–109, Portland, Oregon, USA, 23 June 2011. We used Zettair to generate the ranked list of documents retrieved in response to each query. This work consists in investigating the impact of Multiword Expressions on applications, focusing on compound nouns in Information Retrieval systems, and whether a more adequate treatment for these expressions can bring possible improvements in the indexing these expressions.

Conclusion and Future works

- We have proposed a multi-objective optimization-based approach for summarizing scientific articles where various summary quality measures are simultaneously optimized.
- For the purpose of optimization, the binary differential evolution strategy is considered, which is a meta-heuristic optimization strategy.
- Experiments performed using two computational linguistic datasets prove that our proposed approach is able to outperform the state-of-the-art algorithms.
- An ablation study is also performed by varying the objective functions in our optimization strategy
- In terms of the most contributing features or objective functions for summary generation, sentence position in the article and sentence's semantic similarity with the title have shown good improvement.
- Note that our developed framework is generalized in nature and can be adopted for developing any other summarization systems, including single document summarization, multi-document summarization, microblog summarization, among others. We will be working for developing these systems in future.

Some Relevant References

[1] M. Yasunaga, J. Kasai, R. Zhang, A. R. Fabbri, I. Li, D. Friedman, and D. R. Radev, “Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7386–7393.

[2] Cohan, F. Dernoncourt, D. S. Kim, T. Bui, S. Kim, W. Chang, and N. Goharian, “A discourse-aware attention model for abstractive summarization of long documents,” arXiv preprint arXiv:1804.05685, 2018.

[3] M. Hernandez-Alvarez and J. M. Gomez, “Survey about citation context ´ analysis: Tasks, techniques, and resources,” Natural Language Engineering, vol. 22, no. 3, pp. 327–349, 2016.

[4] K. Jaidka, M. Chandrasekaran, D. Jain, and M.-Y. Kan, “The cl-scisumm shared task 2017: Results and key insights,” 11 2017.

[5] Cohan and N. Goharian, “Scientific document summarization via citation contextualization and scientific discourse,” International Journal on Digital Libraries, vol. 19, no. 2-3, pp. 287–303, 2018.

[6] K. Jaidka, M. K. Chandrasekaran, S. Rustagi, and M.-Y. Kan, “Insights from cl-scisumm 2016: the faceted scientific document summarization shared task,” International Journal on Digital Libraries, vol. 19, no. 2-3, pp. 163–171, 2018

[7] N. Saini, S. Saha, D. Chakraborty, and P. Bhattacharyya, “Extractive single document summarization using binary differential evolution: Optimization of different sentence quality measures,” PloS one, vol. 14, no. 11, 2019.

Thank You!!

Any Queries???

E-Mail: nsaini1988@gmail.com