A SURVEY ON NATURE-INSPIRED OTPIMIZATION ALGORITHMS FOR TEXT SUMMARIZATION

[1]Sushama Dhumal, [2]Dr. Prashant Sharma

[1] Research Scholar, [2] Associate Professor & HOD (CSE),

[1,2] CSE,Pacific University,Udaipur, INDIA

Abstract: Accessibility of immensely available information in electronic form has opened the doors for a newly growing research area of Text Summarization. Text summarization is the process of abbreviating a huge quantity of information in one or more documents into a succinct summary that can contain informative text relevant to the user. Effective summary can be generated with organized information, with the help of optimization algorithms such as particle swarm optimization, artificial bee colony algorithms, genetic algorithms and ant colony optimization are used. These algorithms can be applied on single as well as multiple documents to get optimized summary. This paper aims to present a survey of text summarization procedure using nature-inspired optimization algorithms.

    Index Terms **- Text Summarization; Optimization; Nature-inspired.**

I. INTRODUCTION

Text Summarization is a technique of extracting the core information from the given document and generating a summarized text as a result. Abstractive multi-document summarization generates a concentrated version of the information from a document. Certain optimization techniques can be applied to generate a good quality summarized output. This paper aims to provide the most recent and relevant information in the field of the text summarization using optimization algorithms that could be useful for study and analysis in future research.

Nature-inspired optimization algorithms is a very important research area in computational intelligence, soft computing, and optimization domains in a general sense. Optimization problems in the real world may be with single objective or multiple objectives. Based on the nature of the problem, the algorithms can be modified to multi-objective optimization algorithms.

Swarm Intelligence (SI) is the collective intelligence behaviour of self-organized and decentralized systems. This paper highlights on several Swarm Intelligence Algorithms. Various SI-based algorithms include Genetic Algorithms (GA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Differential Evolution (DE), Artificial Bee Colony (ABC), Glow-worm Swarm Optimization (GSO), and Cuckoo Search Algorithm (CSA) [28]. All of these algorithms have proved themselves in solving many optimization problems.

II. LITERATURE REVIEW

(Mosa et al, 2019) proposed new model to be adequate for achieving many objectives and to satisfy the real-time needs. This survey highlights, for the first time, how the swarm intelligence (SI) optimization techniques are performed to solve the text summarization task efficiently. A convincing justification of Ant Colony Optimization (ACO), has been presented in this paper [1].

(Abualigah et al, 2018) proposes a novel feature selection method using the particle swarm optimization (PSO) algorithm (FSPSOTC) to solve the feature selection problem by creating a new subset of informative text features to improve the performance of the text clustering technique and reduce the computational time. This proposed method is compared with the other well-known algorithms such as genetic algorithm and harmony search algorithm [3].

(Hemeida et al, 2020), in his paper addresses about the implementation of evolutionary optimization algorithms (EOAs) for mining two famous data sets in machine learning by implementing four different optimization techniques, namely, whale optimization algorithm (WOA), dragonfly algorithm (DA), multiverse optimization (MVA), and grey wolf optimization (GWO). The comparative study proves that, the GWO, and MVO provide accurate results over both WO, and DA in terms of convergence, runtime, classification rate, and MSE as measures [4].

(Verma, Pradeepika et al, 2019) In his paper, a novel multi-document summarization scheme based on metaheuristic optimization is introduced that generates a summary by extracting salient and relevant sentences from a collection of documents. The proposed work generates optimal combinations of sentence scoring methods and their respective optimal weights to extract the sentences with the help of a metaheuristic approach known as teaching–learning-based optimization [15].

(Abd Elfattah et al,2019) has introduced a hybrid approach based on the merger of Salp swarm algorithm and the chaos theory. The proposed hybrid approach has been used to evaluate their ability and precision in the clustering process. It is revealed how Salp can operate to find automatically [6].

(Dalal et. al., 2017) presented a bio-inspired text summarization approach based on semantic graph of input document for Hindi text. The proposed approach uses PSO technique to select the summary sentences optimally resulting in reasonably good performance [22].

(Onan et al, 2016) in his paper proposed ensemble method based on static classifier selection involving majority voting error and forward search, as well as a multi-objective differential evolution algorithm. Based on the static classifier selection scheme, our proposed ensemble method incorporates Bayesian logistic regression, naïve Bayes, linear discriminant analysis, logistic regression, and support vector machines as base learners. This paper analyses that multi-objective weighted voting ensemble classifier can predict better than conventional ensemble learning methods such as AdaBoost, bagging, random subspace, and majority voting [8].

(Jagatheeshkumar et al, 2017) proposes in his article, a text clustering algorithm, which is a combination of k-means and Lion Optimization Algorithm (kLOA). K-means is best clustering algorithm, and LOA is to select the initial cluster point and the clusters are formed. This enhanced k-means algorithm reults in the formation of high-quality text clusters [11].

(Sanchez-Gomez et al, 2018) has proposed and designed a Multi-Objective Artificial Bee Colony (MOABC) algorithm for multi-objective approaches. This algorithm is implemented on DUC datasets and has proven improvement with respect to the best and robust results available in the scientific literature [12].

(MirShojaee et al, 2015) to deal with a large volume of documents for summarization, meta-heuristic techniques are the best solution. This paper has applied Cuckoo Search Optimization Algorithm (CSOA) to improve performance of extractive-based summarization method. The obtained results indicate better performance of proposed method compared with other similar techniques. [23].

(Kyoomarsi et al, 2008) In this paper we first analyze some state of the art methods for text summarization. We discuss what the main disadvantages of these methods are and then propose a new method using fuzzy logic. Comparisons of results show that our method beats most methods which use machine learning as their core [25].

(Saini, Saha et al, 2019a) proposed an approach to build some extractive single document text summarization systems using multi-objective optimization. Three techniques proposed are: (1) first is an integration of self-organizing map and multi-objective differential evolution, (2) second is based on multi-objective grey wolf optimizer, and (3) third is based on multi-objective water cycle algorithm. The sentences present in the document are first clustered utilizing the concept of multi-objective clustering. Two objective functions measuring compactness and separation of the sentence clusters in two different ways are optimized simultaneously using MOO framework. The proposed approach is able to automatically detect the number of sentence clusters present in a document and then representative sentences are selected from different clusters using some sentence scoring features to generate the summary [18].

(Al-Abdallah et al, 2017) proposed the use of Particle Swarm Optimization (PSO) algorithm for the extraction of summaries for single Arabic documents [17].

Sriparna Saha(2019) in her paper has briefly described different types of summarization systems available like document summarization, microblog summarization, multimodal summarization, figure-assisted summarization. It also contains comparison with some recent and older techniques developed in terms of Rouge measures. The author also has mentioned few research goals from text summarization, in the coming years [19].

(Asgari et al,2014) in his paper proposed a new method based on multi-agent particle swarm optimization for automatic text summarization. It was tested with a set of DUC 2002 standard documents and the results were analyzed by ROUGE evaluation software and were also compared with other methods [21].

(Dewi et al, 2019) this paper describes about optimizing the Relevance Vector Machine (RVM) algorithm in automatic text summarization to find out what features are commonly used in the summarization process. Each feature value will be calculated as a correlation with target. The composition of features is determined by obtained correlation value, when correlation value between features and targets is grater, the feature will take precedence. This paper concludes that correlation coefficient can be used to determine the order of extraction features [24].

(Saini et al, 2019b) paper proposed two approaches for sentence based extractive single document summarization to utilize a multi-objective based differential evolution technique for optimization. Self-organizing map (SOM) based genetic operators are incorporated in the optimization process to improve the convergence. Best performance depends on objective functions and the correct choice of similarity/dissimilarity measure between sentences [26].

(MirShojaee et al, 2017) this paper summarizes text through objective extraction of some parts of a text like word, sentence, and paragraph. This paper describes a new metaheuristic method which is a biogeography - based optimization algorithm (BBO) for extractive text summarization [27].

III. TYPES OF SUMMARY AND OPTIMIZATION

Several papers for text summarization with optimization referred in this article have summarized either single document or multiple documents, with extractive or abstractive summarization technique. Different SI-based optimization approaches can be applied for summarization.

Table-I briefs about the types of optimization approaches produced by the surveyed literature for text summarization.

Table 3.1 Types of Optimization Approaches

Reference	Types of summaries	Optimization approach
Mosa, et al, 2019	Short text summaries	Swarm intelligence techniques such as Ant Colony Optimization
Abualigah et al, 2018	Document clustering	Particle Swarm Optimization (PSO)
Hemeida et al, 2020	Data mining	whale optimization algorithm (WOA), dragonfly algorithm (DA), multiverse optimization (MVA), and grey wolf optimization (GWO)
Abd Elfattah et al,2019	Document Binarization	Salp swarm algorithm and the chaos theory
Onan et al, 2016	sentiment classification	Multi-objective optimization
Jagatheeshkumar et al, 2017	Textual Data Clustering	Lion Optimization algorithm
Sanchez-Gomez et al, 2018	Extractive multi-document text summarization	multi-objective artificial bee colony optimization
Verma, Pradeepika et al, 2019	text summarization	metaheuristic optimization
Rajkumar et al, 2016	Survey of text summarization with optimization	
Al-Abdallah et al, 2017	Extractive Single document summarization	Particle Swarm Optimization (PSO)
Saini et al, 2019 [a]	Extractive Single document summarization	multi-objective grey wolf optimizer
Sriparna Saha, 2019	Extractive text summarization	Multi-objective optimization
Alguliyev et al, 2018	text summarization	clustering and COSUM optimization
Asgari et al, 2014	extractive text summarization	multi-agent particle swarm optimization
Vipul Dalal et al, 2017	automatic text summarization	particle swarm optimization and semantic graph
Mirshojaei et al, 2015	extractive-based text summarization	Cuckoo Search Optimization Algorithm (CSOA)
Dewi et al, 2019	automatic text summarization	Relevance Vector Machine Optimization
Kyoomarsi et al, 2008	Text summarization	Fuzzy logic
Saini et al, 2019 [b]	Extractive Single document summarization	Multi-objective binary differential evolution
MirShojaee et al, 2017	Text summarization	Cuckoo Search Optimization

IV. EVALUATION CORPORA

Summarization algorithms have been applied on variety of corpora. Evaluation corpus which contains minimum 3000 document is considered as proper to increase the strength of features. Most of the surveyed papers have used DUC (Document Understanding Conference) dataset which is the de-facto standard datasets for NLP community. Some of the surveyed papers have used text documents for evaluation or have prepared their own corpora.

Table-II briefs about the types of corpora used by the surveyed literature.

Table 4.1 Types of corpus used by the surveyed literature

Reference	Corpus Used
Mosa, et al, 2019	Authors' corpus
Abualigah et al, 2018	Text clustering benchmark standard datasets are available at the Laboratory of Computational Intelligence (LABIC)
Hemeida et al, 2020	Iris dataset and Breast Cancer dataset
Abd Elfattah et al,2019	H-DIBCO 2016 dataset containing 10 different handwritten documents images
Onan et al, 2016	public sentiment analysis datasets from different domains, named camera, camp, doctor, drug, laptop, lawyer, radio, TV, and music
Jagatheeshkumar et al, 2017	Textual documents
Sanchez-Gomez et al, 2018	DUC dataset
Verma, Pradeepika et al, 2019	DUC2006, DUC2007
Rajkumar et al, 2016	- NA -
Al-Abdallah et al, 2017	Essex Arabic Corpus (EASC)
Saini et al, 2019 [a]	DUC2001, and DUC2002
Sriparna Saha, 2019	DUC2001, and DUC2002
Alguliyev et al, 2018	DUC2001, and DUC2002
Asgari et al, 2014	DUC 2002
Vipul Dalal et al, 2017	80 Hindi documents
Mirshojaei et al, 2015	Doc 2002 standard
Dewi et al, 2019	Sentences in text document
Kyoomarsi et al, 2008	Text documents
Saini et al, 2019 [b]	DUC2001, and DUC2002
MirShojaee et al, 2017	DUC2002

V. EVALUATION METHODS

To evaluate the quality and the goodness of an extracted summary as well as to compare the results of various summarization systems, appropriate assessment methods are required. There are many numbers of methods used for summary evaluation like Pyramid method, ROUGE [79], etc.

This segment indicates the various evaluation strategies that have been used in this survey paper. Text quality can be measured with two approaches namely, intrinsic and extrinsic. Intrinsic evaluation approach mainly focuses on coherence and informativeness of summaries. Whereas, extrinsic evaluation approach measures the efficiency and acceptability of the generated summaries in some task [29]. It is found that all the surveyed papers have used intrinsic evaluation approach, which is based on text quality, co-selection or sentence extraction and contents.

VI. CONCLUSION

Many researchers have tried to propose different techniques to bring improvement in the summarization of the text documentation. Text summary can be either of two types, extractive and abstractive. The extractive summary is well investigated and plenty of researchers have proposed various summarization structures. The abstractive summary is a complicated challenge and also requires an extractive summary to narrow an abstractive summary. For this reason, many researchers have used extractive approach. Optimization improves quality of the generated summary. This combination has a vast scope for research.

The purpose of this study is to provide a starting point for new researchers by reviewing the categories of text summarization with optimization algorithms. This study also examines the different approaches proposed by different researchers. This paper also has mentioned the important points that need to be considered during the evaluation.

In the text summarization process, some basic steps are common almost across all the techniques. The recent papers have added information about more refined version of this process such as, the part of speech tagging (POS tags) etc. Similarly, Swarm intelligence optimization techniques such as PSO have used by most of the researchers. Different optimization techniques produce different results, but their performance cannot be compared with each other, as each technique proves itself better in one or the other way [1].

Novel approaches can be developed using the successful systems in this domain of research. The evaluation of the summarization is still an interesting research area. Likewise, the various ways of representing the optimized summary can be found out.

REFERENCES

[1] Ali, A. 2001.Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. Journal of Empirical finance, 5(3): 221–240.

[1] Mosa, Mohamed Atef, Arshad Syed Anwar, and Alaa Hamouda. 2019. "A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms." Knowledge-Based Systems 163: 518-532.

[2] Ozbay, Feyza Altunbey, and Bilal Alatas. 2019. "A Novel Approach for Detection of Fake News on Social Media Using Metaheuristic Optimization lgorithms", Elektronika ir Elektrotechnika 25.4: 62-67.

[3] Abualigah, Laith Mohammad, Ahamad Tajudin Khader, and Essam Said Hanandeh. 2018. "A new feature selection method to improve the document clustering using particle swarm optimization algorithm." Journal of Computational Science 25: 456-466.

[4] Hemeida, A. M., et al. 2020. "Implementation of nature-inspired optimization algorithms in some data mining tasks." Ain Shams Engineering Journal 11.2: 309-318.

[5] S. Radha Priya , Dr. M. Devapriya, 2020. "Twitter Sentiment Analysis with Diabetic Drugs Using Machine Learning Techniques with Glowworm Swarm Optimization Algorithm", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 07.

[6] Abd Elfattah, Mohamed, Aboul Ella Hassanien, and Sherihan Abuelenin. 2019. "A hybrid swarm optimization approach for document binarization." Studies in Informatics and Control 28.1: 65-76.

[7] Rodríguez, Nibaldo, et al. 2018."Optimization algorithms combining (meta) heuristics and mathematical programming and its application in engineering." Mathematical Problems in Engineering 2018.

[8] Onan, Aytuğ, Serdar Korukoğlu, and Hasan Bulut. 2016. "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification." Expert Systems with Applications 62: 1-16.

[9] Ahmadian, Ali, Ali Elkamel, and Abdelkader Mazouz. 2019. "An improved hybrid particle swarm optimization and tabu search algorithm for expansion planning of large dimension electric distribution network." Energies 12.16: 3052.

[10] Singh, J., Singh, G. & Singh, R. 2017. "Optimization of sentiment analysis using machine learning classifiers". Hum. Cent. Comput. Inf. Sci. 7, 32.

[11] Jagatheeshkumar, G., and S. SELVA BRUNDA. 2017. "Lion Optimization Algorithm Based K-Means For Textual Data Clustering." International Journal of Pure and Applied Mathematics 117.22: 167-171.

[12] Sanchez-Gomez, Jesus M., Miguel A. Vega-Rodríguez, and Carlos J. Pérez. 2018. "Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach." Knowledge-Based Systems 159: 1-8.

[13] Valdez, Fevrier, et al. 2019. "Nature-inspired optimization algorithms for neuro-fuzzy models in real-world control and robotics applications.".

[14] Shankar, K., and S. K. Lakshmanaprabu. 2018. "Optimal key based homomorphic encryption for color image security aid of ant lion optimization algorithm." International Journal of Engineering & Technology 7.9: 22-27.

[15] Verma, Pradeepika, and Hari Om. 2019. "A novel approach for text summarization using optimal combination of sentence scoring methods." Sādhanā 44.5: 110.

[16] Rajkumar, V.C.Thilak. 2016. "A Survey On Text Summarization Using Optimization Algorithm.", ELK Asia Pacific Journal Of Computer Science And Information Systems, ISSN: 2454-3047; ISSN: 2394-0441 (Online) Volume 2 Issue 1.

[17] Al-Abdallah, Raed Z., and Ahmad T. Al-Taani. 2017. "Arabic single-document text summarization using particle swarm optimization algorithm." Procedia Computer Science 117: 30-37.

[18] Saini, Naveen, et al. 2019. "Extractive single document summarization using multi-objective optimization: Exploring self-organized differential evolution, grey wolf optimizer and water cycle algorithm." Knowledge-Based Systems 164: 45-67.

[19] Saha, Sriparna. 2019. "Text summarization using multi-objective optimization." CSI Transactions on ICT 7.3: 251-255.

[20] Alguliyev, Rasim M., et al. 2019. "COSUM: Text summarization based on clustering and optimization." Expert Systems 36.1 (2019): e12340.

[21] Asgari, Hamed, Behrooz Masoumi, and Omid Sojoodi Sheijani. 2014. "Automatic text summarization based on multi-agent particle swarm optimization." 2014 Iranian Conference on Intelligent Systems (ICIS). IEEE.

[22] Vipul Dalal , Dr. Latesh Malik. 2017. "Automatic Summarization for Hindi Text Documents using Bio-inspired Computing", International Journal of Advanced Research in Computer and Communication Engineering ,Vol. 6, Issue 4, April 2017: ISSN (Online) 2278-1021.

[23] Mirshojaei, Seyed Hossein, and Behrooz Masoomi. 2015. "Text summarization using cuckoo search optimization algorithm." Journal of Computer & Robotics 8.2: 19-24.

[24] Dewi, K. E., and E. Rainarli. 2019. "Relevance Vector Machine Optimization in Automatic Text Summarization." IOP Conference Series: Materials Science and Engineering. Vol. 662. No. 5. IOP Publishing.

[25] Kyoomarsi, Farshad, et al. 2008. "Optimizing text summarization based on fuzzy logic." Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008). IEEE.

[26] Saini, Naveen, et al. 2019. "Extractive single document summarization using binary differential evolution: Optimization of different sentence quality measures." PloS one 14.11: e0223477.

[27] MirShojaee, Hossein, Behrooz Masoumi, and Esmaeel Zeinali. 2017. "Biogeography-based optimization algorithm for automatic extractive text summarization." International Journal of Industrial Engineering & Production Research 28.1: 75-84.

[28] Ab Wahab MN, Nefti-Meziani S, Atyabi A. 2015. "A Comprehensive Review of Swarm Optimization Algorithms", PLoS ONE 10(5): e0122827. doi:10.137.

[29] M Indu, Kavitha K V. 2016. "Review on text summarization evaluation methods", International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07.