Let's Summarize Scientific Documents! A Clustering-based Approach via Citation Context

Santosh Kumar Mishra[1], Naveen Saini[2], Sriparna Saha[1], Pushpak Bhattacharyya[3]
Email: {santosh_1821cs03, sriparna}@iitp.ac.in[1], naveensaini@wsu.ac.kr[2], pb@cse.iitb.ac.in[3]

[1] Indian Institute of Technology Patna, Patna, India 801103
[2] Technology Studies Department, Endicott College of International Studies, Woosong University, South Korea
[3] Indian Institute of Technology Bombay, Mumbai, India 400076

Abstract. Scientific documents are getting published at expanding rates and create challenges for the researchers to keep themselves up to date with the new developments. Scientific document summarization solves this problem by providing summaries of essential facts and findings. We propose a novel extractive summarization technique for generating a summary of scientific documents after considering the citation context. The proposed method extracts the scientific document's relevant sentences with respect to citation text in semantic space by utilizing the word mover's distance (WMD); further, it clusters the extracted sentences. Moreover, it assigns a rank to cluster of sentences based on different aspects like similarity with the title of the paper, position of the sentence, length of the sentence, and maximum marginal relevance. Finally, sentences are selected from different clusters based on their ranks to form the summary. We conduct our experiments on CL-SciSumm 2016 and CL-SciSumm 2017 data sets. The obtained results are compared with the state-of-the-art techniques. Evaluation results show that our method outperforms others in terms of ROUGE-2, ROUGE-3, and ROUGE-SU4 scores.

Keywords: Scientific Summarization · Clustering · Word Mover's Distance · Maximum Marginal Relevance

1 Introduction

The publication rate of scientific papers is increasing day by day; the availability of the massive amount of scientific literature is a big challenge for researchers in various fields to keep them up-to-date with the new developments. A recent study by bibliometric analysts shows that global scientific output doubles after every nine years [2]. Scientific document summarization aims to solve this problem by summarizing the important contributions and findings of the reference paper [7] [5] [6] and thus, reducing the effort of the researchers to understand the paper. There are two approaches to scientific summarization in the literature; the first is the abstract of the document. Though the paper's abstract provides the paper's theme, but may not convey the all-important contributions and impact of the paper. The same has also been shown in recent paper [18][1]. These kinds of problems motivate the researcher to solve the scientific summarization task using the second approach, i.e., citation-based summarization [16] [8] [7]. Citation based summary is obtained by utilizing a set of citations referring to the original document. Citations are a short description that explains the proposed method, result, and important findings of the cited work; this description is known as citation text or citance.

This paper proposes a novel approach for scientific document summarization using an extractive summarization technique that extracts important sentences from the reference paper. Here, we extract important sentences for each citation of the reference paper-based on semantic similarity between citation text and sentences of the reference paper using word mover's distance [13]. Further, we apply clustering on all distinct important sentences. Then, we rank the clusters based on the distances between the cluster center and the document center (representative sentence of the document). Finally, we extract sentences from ranked clusters using several sentence-scoring features until the summary's desired length (i.e., 250 words) is reached. The proposed approach is evaluated on two datasets: CL-SciSumm 2016 and CL-SciSumm 2017, related to the computational linguistic domain.

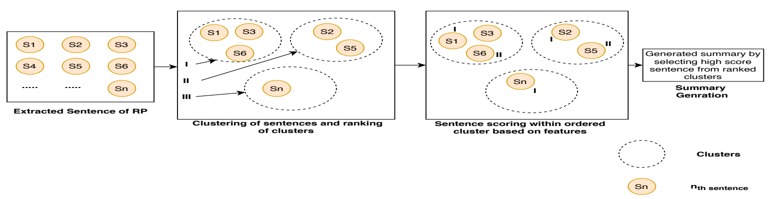

Fig. 1. Process flow chart of proposed method

2 Proposed Methodology

In this section, the steps followed in our proposed framework are discussed. The flowchart of the proposed approach is shown in Figure 1.

2.1 Extracting the Citation Context

Initially, we have extracted the sentences from the reference paper, which is to be summarized. For this purpose, we have utilized the word mover's distance. Here, we have computed the word mover's distance (WMD) between each citation sentence and the reference paper's sentences. Then the top five sentences which are having minimum WMD are selected [11]. Let the set of distinct important sentences extracted from the reference paper (RP) after considering all citations be denoted by $\mathcal{S}$. Note that WMD calculates the similarity between the sentences in terms of distance [13], where minimum distance represents more similarity between sentences.

2.2 Grouping of Sentences using Clustering

Sentences in $\mathcal{S}$ obtained in the previous step are grouped using the K-Medoids [19] clustering algorithm. It utilizes WMD as a distance measure between sentences instead of Euclidean sentence and, thus, is able to capture the semantic similarity present between the sentences. We have used the K-medoid clustering with the number of clusters decided by the elbow method. Let the obtained cluster centers be represented as $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K\}$.

2.3 Ranking the Clusters Obtained

It includes two steps: representative sentence calculation and ranking of clusters, which are discussed below:

Representative Sentence Calculation: After getting clusters of sentences, it is required to build a summary. But, it is very difficult to decide which cluster should be considered first to extract the sentences. Thus, there is a need to rank the clusters. Therefore, to perform the same, firstly, we have determined the document center/representative sentence (RP) of the document. It is that sentence in the document which is the most similar to the remaining sentences. We can also call it as an document's center. In other words, among $\mathcal{S}$, the sentence having the minimum average WMD with respect to other sentences is called the RP. Mathematically, it is defined as $r = argmin \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \frac{wmd(s_i, s_j)}{M}$

Here, r is the index of representative sentence in $\mathcal{S}$, N is the total number of distinct sentences in set $\mathcal{S}$ and M is the number of sentence pairs, equals to $\frac{N*N-1}{2}$. s_i and s_j are the i^{th} and j^{th} sentence in the set $\mathcal{S}$, respectively.

Ranking of Clusters: To rank the clusters, WMD distance between the cluster center, $\mathcal{C}_i$ $(1 \leq i \leq K)$, and representative sentence, r, is calculated. Clusters are ranked based on their distances from the representative sentence means. The cluster closest to the representative sentence is assigned the highest priority. $d_i = wmd(\mathcal{C}_i, \mathcal{S}_r) \quad \forall i \in 1, 2, \ldots, K$

Here, d_i denotes the distance between i^{th} cluster center and representative sentence, $\mathcal{S}_r$. Then, these distances are sorted in ascending order. The cluster, which is at the lowest distance, is assigned rank-1 and so on. In other words, sentences are extracted from the higher rank to the lower rank clusters.

2.4 Calculating Sentence Scores in Each Cluster

After assigning ranks to different clusters, sentence scores are calculated in each cluster using different aspects/features. These scores help in selecting sentences from a cluster that will be part of the summary. These features are described below:

Similarity with Paper's Title (F_1): WMD between the title of the document and the sentences of the cluster has been calculated. The sentence is given the highest priority, which has minimum WMD distance with respect to the title [17].

Position of the Sentence (F_2): In most of the documents or papers, important sentences are found in the title and lead sentences of a paragraph; it is expressed as follows $m_i = \sqrt{\frac{1}{n_i}}$ where n_i is the position of a sentence in the reference paper. The sentence is given the highest priority, which lies at the starting of the paper or document [17].

Length of the Sentence (F_3): In the literature, it is shown that the longest sentences of the document are always relevant for the summaries [15, 17]. The sentence is assigned the highest priority, which has the longest length.

Maximum Marginal Relevance (F_4): This feature is used to maintain anti-redundancy in the summary [3]. Sentences from each cluster are selected based on the following formula: $score(X) = \lambda Sim_1(s, D) - (1 - \lambda) Sim_2(s, Summary)$

Here, score(X) represents linear interpolation of Sim_1 and Sim_2 where Sim_1 is the similarity of a sentence with respect to all other sentences in the cluster, and Sim_2 is the similarity of a sentence with respect to the sentences that are already included in the summary, D is the document (extracted sentences using citation context), and s is the sentence that is going to be included in the summary.

We have used WMD for the similarity between sentences. Here, $\lambda = 0.7$ which is used in [4] . The sentence is assigned the highest priority, which has the highest score of X.

2.5 Summary Generation

For the summary generation, we have considered the clusters in a rank-wise manner. Given the clusters, the summary is generated by selecting the highest ranked sentence from each cluster based on the above four features. We have generated a summary utilizing each feature and evaluated it against different types of summaries available with the datasets.

3 Experimental Setup

3.1 Datasets Used

In the current paper, we have utilized two datasets, namely, CL-SciSumm 2016 and CL-SciSumm 2017, to evaluate our method. Details of the datasets can be found at https://github.com/WING-NUS/scisumm-corpus.

3.2 Evaluation Metrics

The proposed method is evaluated with well-known evaluation metric, ROUGE score [14] for evaluating the summarization outputs.

3.3 Comparative methods

We have compared the proposed method with the state-of-the methods of CL-SciSumm 2016 and CL-SciSumm 2017, these methods can be found in [11] and [10], respectively.

4 Results and Discussions

4.1 Results with CL-SciSumm 2016 Dataset

The results of the proposed method on the CL-SciSumm 2016 data set are shown in Table 1. This table is divided into two parts: (a) results of the proposed method using different features, (b) best results as compared with the state-of-the-art systems [11] of CL-SciSumm 2016. From these Tables it can be concluded that, the proposed method has better scores for the human summary and community summary, whereas, for the abstract summary, it lacks behind by only one system, namely, $sys8PARA7$. For the human summary, feature F2 is the most contributing feature. The proposed method has attained the highest ROUGE-SU4 score of 0.190, whereas the highest score reported in existing methods is 0.136. Our proposed approach has attained 39.70% improvement in terms of the ROUGE-SU4 score. For community summary, also, F2 is the most important feature, and our proposed approach has attained the highest ROUGE-SU4 score of 0.240, whereas the highest score reported in existing systems for CL-SciSumm 16 dataset is 0.167. Our method has obtained 43.71% improvements in terms of ROUGE-SU4. For the abstract summary, feature F3 is the best performing feature. Our proposed approach has attained a ROUGE-SU4 score of 0.308, which is the second-highest score after $sys8PARA7$.

Results of the proposed method are compared with some recent systems developed by Cohan et al. [6]; the corresponding results are shown in Table 2. It can be concluded from the table that our method performs better in terms of ROUGE-2 and ROUGE-3 scores except for one supervised model; our approach is unsupervised; this can be a reason behind the second-best performance. Note that Cohan et al. [6] have used citation contextualization and discourse facet. Our method does not use discourse facet as it needs supervised learning; our method is purely unsupervised in nature.

4.2 Results with CL-SciSumm 2017

The results of the proposed method on the CL-SciSumm 2017 dataset are shown in Table 3. Similar to Table 1 and Table 2, this table also consists of two parts: (a) results obtained using various features; (b) best results compared with the state-of-the-art system (methods) of CL-SciSumm 2017 [9]. It can be concluded from Table 3 that our proposed method performs better than all other systems for the community summary. For human summary, our proposed method has attained the highest ROUGE-SU4 score of 0.234 with 31.46% improvements over the best existing system, whereas, in terms of ROUGE-2 score, our method has attained less score in comparison to some of the systems. For community summary, our method has attained highest scores in terms of ROUGE-2 and ROUGE-SU4 metrics, which are 15.68% and 59.19% improvements in terms of ROUGE-2 and ROUGE-SU4 scores, respectively. For the abstract summary, our method has attained a better score than many methods in terms of ROUGE-2 and ROUGE-SU4 scores, but those are not the best ones. Note that the abstract is written by human authors, and our system is based on extractive summarization; therefore, this could be the reason behind poor performance by the proposed method.

Table 1. (a). Scores of generated summary in terms of ROUGE-SU4 against human summary, community Summary and abstract. (b) Comparison of performance of our proposed method with respect to state-of-the-art methods reported in CL-SciSumm 16 [12] in terms of ROUGE-SU4 metric. Here HS denotes human-summary, CS denotes community-summary and Abs denotes abstract.

Methods	HS	CS	Abs
F1	0.139	0.201	0.193
F2	**0.190**	**0.240**	0.304
F3	0.108	0.171	0.115
F4	0.176	0.228	**0.308**

(a)

Methods	HS	CS	Abs
sys8$PARA_7	0.136	0.130	0.423
sys3$LMKL1_CCS1	0.124	0.095	0.179
sys3$LMEQAL_CCS2	0.121	0.102	0.214
sys3$LMKL2_CCS3	0.114	0.095	0.158
sys8$PARA_1	0.112	0.129	0.247
sys8$PARA_8	0.111	0.150	0.244
sys3$TFCCS4	0.101	0.085	0.129
sys8$PARA_0	0.099	0.137	0.177
sys8$PARA_4	0.094	0.162	0.170
sys10$AUTOMATIC	0.092	0.150	0.124
sys15$TKERN18	0.090	0.096	0.102
sys15$TFIDF+ST+SL	0.088	0.167	0.092
sys15$TKERN14CE	0.085	0.129	0.105
sys10$COMMUNITY	0.085	0.149	0.111
sys15$TKERN11CE	0.082	0.106	0.105
sys15$TKERN11	0.081	0.103	0.107
sys15$TKERN14	0.080	0.110	0.099
sys15$TKERN18CE	0.071	0.103	0.093
sys5$DEFAULT	0.065	0.082	0.087
sys16$DEFAULT	0.048	0.107	0.053
Proposed Method	**0.190**	**0.240**	**0.308**

(b)

Table 2. (a). Scores of generated summary against human summary in terms of ROUGE-2 and ROUGE-3 for CL-SciSumm 16 dataset. (b). Comparison of performance of our method for human summary with respect to state-of-the-art methods reported in [6] in terms of ROUGE-2 and ROUGE-3 scores

Methods	ROUGE	
	ROUGE-2	**ROUGE-3**
F1	0.164	0.116
F2	**0.235**	**0.175**
F3	0.122	0.073
F4	0.220	0.168

(a)

Methods	ROUGE-2	
	ROUGE-2	**ROUGE-3**
BM25	0.152	0.130
VSM	0.148	0.127
LM	0.143	0.126
QR-NP	0.158	0.136
QR-KW	0.160	0.138
$\mathrm{WE}_{w}iki$	0.145	0.125
$\mathrm{WE}_{w}iki + retrofit$	0.147	0.137
Supervised	0.175	0.150
Proposed Method	**0.235**	**0.175**

(b)

Table 3. Scores of generated summary in terms of ROUGE-2 and ROUGE-SU4 against human summary, Community Summary and Abstract. (b) Comparison of our proposed method with respect to state-of-the-art methods reported in CL-SciSumm 17 [9] in term of ROUGE-2 (R-2) and ROUGE-SU4 (R-SU4) scores; Here HS denotes human-summary, CS denotes community summary and Abs denotes Abstract.

Methods	HS		CS		Abs	
	R-2	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**
F1	0.111	0.177	0.199	**0.267**	**0.146**	**0.197**
F2	**0.153**	**0.234**	**0.219**	0.264	0.108	0.151
F3	0.057	0.135	0.164	0.218	0.057	0.103
F4	0..070	0.121	0.091	0.138	0.080	0.101

(a)

Methods	HS		CS		Abs	
	R-2	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**
CIST Run 4	0.156	0.101	0.184	0.136	0.351	0.185
CIST Run 1	0.171	0.111	0.187	0.137	0.341	0.167
CIST Run 6	0.184	0.110	0.185	0.141	0.331	0.172
CIST Run 3	0.275	0.178	0.204	0.168	0.327	0.171
CIST Run 2	0.225	0.147	0.195	0.155	0.322	0.163
CIST Run 5	0.153	0.118	0.192	0.146	0.318	0.178
UPF summa_abs	0.168	0.147	0.190	0.153	0.297	0.158
UPF acl_abs	0.214	0.161	0.191	0.167	0.289	0.163
UniMa Runs 1, 2, 3	0.197	0.157	0.181	0.169	0.265	0.184
NJUST Run 4	0.206	0.131	0.167	0.126	0.258	0.152
UniMa run 4, 5, 6	0.221	0.166	0.178	0.174	0.257	0.191
UniMa run 7, 8, 9	0.224	0.169	0.167	0.167	0.256	0.187
UPF summa_com	0.168	0.142	0.178	0.143	0.247	0.153
CIST Run 7	0.170	0.133	0.163	0.141	0.240	0.154
NJUST Run 2	0.229	0.154	0.152	0.114	0.214	0.138
NJUST Run 1	0.190	0.114	0.147	0.101	0.198	0.114
NJUST Run 5	0.178	0.127	0.119	0.098	0.192	0.108
Jadavpur Run1	0.181	0.129	0.132	0.119	0.191	0.133
NJUST Run 3	0.162	0.115	0.141	0.127	0.187	0.119
UPF google_abs	0.172	0.132	0.143	0.139	0.170	0.108
UPF acl_com	0.217	0.166	0.189	0.169	0.161	0.099
UPF summa_hum	0.189	0.148	0.131	0.147	0.144	0.091
UPF acl_hum	0.188	0.147	0.132	0.127	0.124	0.102
UPF google_hum	0.127	0.101	0.103	0.109	0.071	0.071
UPF google_com	0.120	0.092	0.075	0.096	0.052	0.065
Proposed Method	**0.153**	**0.234**	**0.219**	**0.267**	**0.146**	**0.197**

(b)

4.3 Ranked Analysis of the Results

It can be concluded from the previous sections, for CL-SciSumm 2016 and CL-SciSumm 2017 datasets, no system (Table 1 and Table 3) is the best suited for the human summary, community summary, and abstract summary. It can be seen from Table 1 that system $sys8PARA7$ has the best score for abstract summary (as shown in Table 1), but it is not the best system for human summary and community summary. Similarly, if we observe Table 3, system $CISTRUN4$ is the best system for an abstract summary in terms of ROUGE-2 score, but it is not the best system for human summary and community summary generations. To resolve the ties and analyze the performance of different methods, the ranking based analysis of all methods (systems) proposed in the CL-SciSumm 2016 is shown in Table 4 (a), whereas for CL-SciSumm 2017, the same is illustrated in Table 4 (b). In the ranking table, each method is assigned a rank according to its performance. Each of the systems is assigned a rank value for the human summary, community summary, and abstract summary. Finally, each system

Table 4. Ranking based comparison with state of the art techniques for CL-SciSumm 16 (a) and CL-SciSumm 17 (b); Here AR denotes average ranking.

SOTA	HS	CS	Abs	AR
F2	1	1	1	1.66
F4	2	2	2	2
F1	3	3	7	4.33
sys8$PARA_7	4	12	1	5.66
sys8$PARA_8	9	8	5	7.33
sys8$PARA_1	8	13	4	8.33
F3	10	4	14	9.33
sys8$PARA_4	13	6	10	9.66
sys8$PARA_0	12	10	9	10.33
sys3$LMEQUAL_CCS2	6	19	6	10.33
sys3$LMKL1_CCS1	5	21	8	11.33
sys10$AUTOMATIC	14	7	13	11.33
sys3$LMKL2_CCS3	7	22	11	13.33
sys10$COMMUNITY	17	9	15	13.66
sys15$TFIDF+ST+SL	16	5	22	14.33
sys3$TF_CCS4	11	23	12	15.33
sys15$TKERN14CE	18	22	18	16
sys15$TKERN11CE	19	16	17	17.33
sys15$TKERN11	20	17	16	17.66
sys15$TKERN18	15	20	19	18
sys15$TKERN14	21	14	20	18.33
sys15$TKERN18CE	22	18	21	20.33
sys16$DEFAULT	24	15	24	21
sys5$DEFAULT	23	24	23	23.33

(a)

SOTA	HS		CS		Abs		AR	AR
	R-2	R-SU4	R-2	R-SU4	R-2	R-SU	R-2	R-SU4
F1	27	2	3	2	22	1	17.33	1.66
UNIMA Run 4, 5,6	5	5	13	4	11	2	9.66	3.66
UNIMA Run 7,8,9	4	4	16	9	12	3	10.66	5.33
CIST Run 3	1	3	2	7	4	8	2.33	6
F2	23	1	1	1	25	16	16.33	6
UNIMA Run 1,2,3	9	8	12	5	9	5	10	6
UPF acl_abs	7	7	6	8	8	11	7	8.66
CIST Run 2	3	11	4	10	5	10	4	10.33
UPF summ_abs	19	12	7	11	7	12	11	11.66
UPF acl_com	6	6	8	6	21	26	11.66	12.66
CIST Run 5	24	22	5	13	6	6	11.66	13.66
UPF summ_com	20	14	14	14	13	14	15.66	14
F3	29	15	17	3	28	25	24.66	14.33
CIST Run 7	18	16	18	16	14	13	16.66	15
CIST Run 6	12	26	10	15	3	7	8.33	16
UPF summ_Hum	11	10	25	12	23	27	19.66	16.33
CIST Run 4	22	27	11	20	1	4	11.33	17
NJUST Run 2	2	9	19	25	15	17	12	17
CIST Run 1	16	25	9	19	2	9	9	17.66
NJUST Run 4	8	18	15	23	10	15	11	18.66
UPF google_abs	17	17	21	17	20	23	19.33	19
F4	28	21	28	18	26	20	27.33	19.66
UPF acl_hum	12	13	24	22	24	24	20	19.66
Jadavpur, Run 1	14	19	23	24	18	18	18.33	20.33
NJUST, Run 3	21	23	22	21	19	19	20.66	21
NJUST Run 5	15	20	26	28	17	22	19.33	23.33
NJUST Run 1	10	24	20	27	16	21	15.33	24
UPF google_hum	25	28	27	26	27	28	26.33	27.33
UPF google_com	26	29	29	29	29	29	28	29

(b)

is assigned an average rank, which is the average of the ranks over the human summary, community summary, and abstract.

For CL-SciSumm 2016 and CL-SciSumm 2017 datasets, the ranking Tables are shown in Table 4 (a) and Table 4 (b), respectively. It can be concluded from Table 4 (a) that our proposed method is the best one among all the submitted systems. On the other hand, from Table 4 (b) for CL-SciSumm 2017 dataset, it can be concluded that our proposed method is the best among all the systems in terms of ROUGE-SU4 score. In terms of the ROUGE-2 score, our method is at 17^{th} position in the overall ranking.

5 Conclusion

We present a clustering-based method for scientific document summarization. We utilize word mover's distance to extract the citation context. Incorporating different features like maximal marginal relevance, sentence position in the document, among others, helps in the summary generation process. The obtained results illustrate our proposed method's efficacy over the state-of-the-art techniques in most cases. In future, multi-objective optimization-based clustering can be used for scientific document summarization.

References

1. Atanassova, I., Bertin, M., Larivière, V.: On the composition of scientific abstracts. Journal of Documentation **72**(4), 636–647 (2016)
2. Bornmann, L., Mutz, R.: Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology **66**(11), 2215–2222 (2015)
3. Carbonell, J.G., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: SIGIR. vol. 98, pp. 335–336 (1998)

4. Cohan, A., Goharian, N.: Scientific article summarization using citation-context and article's discourse structure. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 390–400. Association for Computational Linguistics, Lisbon, Portugal (Sep 2015). https://doi.org/10.18653/v1/D15-1045, https://www.aclweb.org/anthology/D15-1045
5. Cohan, A., Goharian, N.: Scientific article summarization using citation-context and article's discourse structure. arXiv preprint arXiv:1704.06619 (2017)
6. Cohan, A., Goharian, N.: Scientific document summarization via citation contextualization and scientific discourse. International Journal on Digital Libraries **19**(2-3), 287–303 (2018)
7. Cohan, A., Soldaini, L., Goharian, N.: Matching citation text and cited spans in biomedical literature: a search-oriented approach. In: proceedings of the 2015 conference of the North American Chapter of the association for computational linguistics: Human language technologies. pp. 1042–1048 (2015)
8. Hernández-Alvarez, M., Gomez, J.M.: Survey about citation context analysis: Tasks, techniques, and resources. Natural Language Engineering **22**(3), 327–349 (2016)
9. Jaidka, K., Chandrasekaran, M., Jain, D., Kan, M.Y.: The cl-scisumm shared task 2017: Results and key insights (11 2017)
10. Jaidka, K., Chandrasekaran, M.K., Jain, D., Kan, M.Y.: The cl-scisumm shared task 2017: Results and key insights. In: BIRNDL@SIGIR (2017)
11. Jaidka, K., Chandrasekaran, M.K., Rustagi, S., Kan, M.Y.: Overview of the cl-scisumm 2016 shared task. In: Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL). pp. 93–102 (2016)
12. Jaidka, K., Chandrasekaran, M.K., Rustagi, S., Kan, M.Y.: Insights from cl-scisumm 2016: the faceted scientific document summarization shared task. International Journal on Digital Libraries **19**(2-3), 163–171 (2018)
13. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In: International conference on machine learning. pp. 957–966 (2015)
14. Lin, C.Y.: ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out. pp. 74–81. Association for Computational Linguistics, Barcelona, Spain (Jul 2004), https://www.aclweb.org/anthology/W04-1013
15. Mendoza, M., Bonilla, S., Noguera, C., Cobos, C., León, E.: Extractive single-document summarization based on genetic operators and guided local search. Expert Systems with Applications **41**(9), 4158–4169 (2014)
16. Qazvinian, V., Radev, D.R., Mohammad, S.M., Dorr, B., Zajic, D., Whidby, M., Moon, T.: Generating extractive summaries of scientific paradigms. Journal of Artificial Intelligence Research **46**, 165–201 (2013)
17. Saini, N., Saha, S., Chakraborty, D., Bhattacharyya, P.: Extractive single document summarization using binary differential evolution: Optimization of different sentence quality measures. PloS one **14**(11) (2019)
18. Yasunaga, M., Kasai, J., Zhang, R., Fabbri, A.R., Li, I., Friedman, D., Radev, D.R.: 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, pp. 7386–7393 (2019)
19. Zhang, Q., Couloigner, I.: A new and efficient k-medoid algorithm for spatial clustering. In: International conference on computational science and its applications. pp. 181–189. Springer (2005)