

CIST@CLSciSumm-18: Methods for Computational Linguistics Scientific Citation Linkage, Facet Classification and Summarization

Lei Li, Junqi Chi, Moye Chen, Zuying Huang, Yingqi Zhu, and Xiangling Fu

Beijing University of Posts and Telecommunications (BUPT)
No.10 Xitucheng Road, Haidian District, Beijing, P.R.China
{leili,cjq,myc,zoehuang,zhuyq, fuxiangling}@bupt.edu.cn

Abstract. Our system makes contributions to the shared Task 1A (citation linkage), Task 1B (facet classification) and Task 2 (summarization) in CLSciSumm-18@SIGIR2018. We develop it based on our former one called CIST@CLSciSumm-17 [7]. We try to improve the methods for all the shared tasks. We adopt Word Mover's Distance (WMD) and improve LDA model to calculate sentence similarity for citation linkage. We try more methods for facet classification. And in order to improve the performance of summarization, we also add WMD sentence similarity to construct new kernel matrix used in Determinantal Point Processes (DPPs).

Keywords: WMD · LDA · DPPs · Random Forest

1 Introduction

With the development of science and network technology, more and more scientific literature appears, especially in Computational Linguistics (CL) domain. We all make literature surveys in our research for a specific topic to obtain inspiration and novel approaches. However, it's time-consuming for human to analyze all the related contents. The goal of CLSciSumm-18 [1] is to explore summarization of scientific research for CL domain, support research in automatic scientific document summarization and provide evaluation resources to push the current state-of-the-art [2].

CLSciSumm-18 contains **Task 1A**, **Task 1B** and **Task 2**. Each topic of the training dataset and test one consists of a Reference Paper (RP) and several Citing Papers (CPs) with citations to the RP. **Task 1A** is to identify the spans of text (cited text spans, CTS) in the RP for each citance given the RP and CPs. And all CTS might be a sentence fragment, a full sentence, or several consecutive sentences (no more than 5). **Task 1B** requires that for each CTS, we need to identify what facet it belongs to from a predefined set of facets (Aim_Citation, Method_Citation, Implication_Citation, Results_Citation and Hypothesis_Citation). We will generate a structured summary of the RP in **Task 2**, in which there are two types: faceted summary of the traditional

self-summary and the community summary (the collection of citation sentences, 'citances').

In this paper we will introduce our methods, strategies and experiments of **Task 1A**, **Task 1B** and **Task 2** based on our former one called CIST@CLSciSumm-17 [7]. We try to apply new sentence similarity computed from WMD and improved LDA (Latent Dirichlet Allocation) model with better topic features for **Task 1A**. In **Task 1B**, we use more classification methods to obtain the facet of CTS. In **Task 2**, we try WMD sentence similarity to construct kernel matrix for improving the quality of Determinantal Point Processes (DPPs) sampling on the basis of our former work on summarization [3].

2 Related Work

Methods of information extraction and content linkage have sprung up recently, which attract the interest of researchers, especially in the last two years. Methods as well as results of CLSciSumm-2016 and CLSciSumm-2017 are described in [4] [5]. The methods demonstrated in **Task 1A** are highly relevant to the methods of calculating similarity. For example, Ma S et al. [6] combine Similarity-based features (LDA/Jaccard/IDF/TF-IDF/Doc2Vec similarity) with Rule-based features to obtain citation linkage. Li L et al. [7] also propose many similarity methods. Zhang D et al. [8] utilize Search-based Similarity Scoring and Supervised Method. The calculation the Cosine Similarity was used in [9]. Aburaed et al. [10] use Voting system to obtain the best result of Word Embeddings Distance system, Modified Jaccard system and BabelNet Embeddings Distance system. Methods based on measuring semantic textual similarity are used in [11]. Besides, other methods are also applied for citation linkage. **Task 1A** was transformed to a query problem in [12]. Different ranking models and query generation strategies were applied in their system. Karimi et al. [13] use the following approaches: structural correspondence learning, positional language models and textual entailment. For **Task 1B**, we treat it as classification problem. So many classification methods are used in **Task 1B**. Classification methods are mainly divided into two parts: Rule-based methods and supervised machine learning methods [6] [7] [13] [11]. Besides, some other methods are also used in **Task 1B**. For example, Felber et al. [12] transform the span of text into a query problem, and then conduct a majority vote on the top five retrieved results to determine the discourse facet. Prasad et al. [14] use classification and ranking method.

As for summary generation in **Task 2**, some teams submitted their results in BIRNDL 2017. Ma S et al. [6] divide the process into two main steps. They group sentences into different clusters by bisecting K-means, and then use maximal marginal relevance (MMR) to extract sentence from each cluster and combine them into a summary. Aburaed et al. [10] score the sentence using multi-features with different weights, and then get the summary according to the score. Li L et al. [7] make a linear combination of multiple features to compute sentence quality. Besides, they also sample sentences based on Jaccard similarity and

sentence quality. We will try new similarity method to construct new kernel matrix of DPPs for better summary.

3 Methods

The framework of our system is shown in Fig. 1. We first obtain the CTS in RP for each citance in CPs, then use features extracted from CTS to determine its facet, and finally we use CTS and its Facet to generate a summary (no more than 250 words).

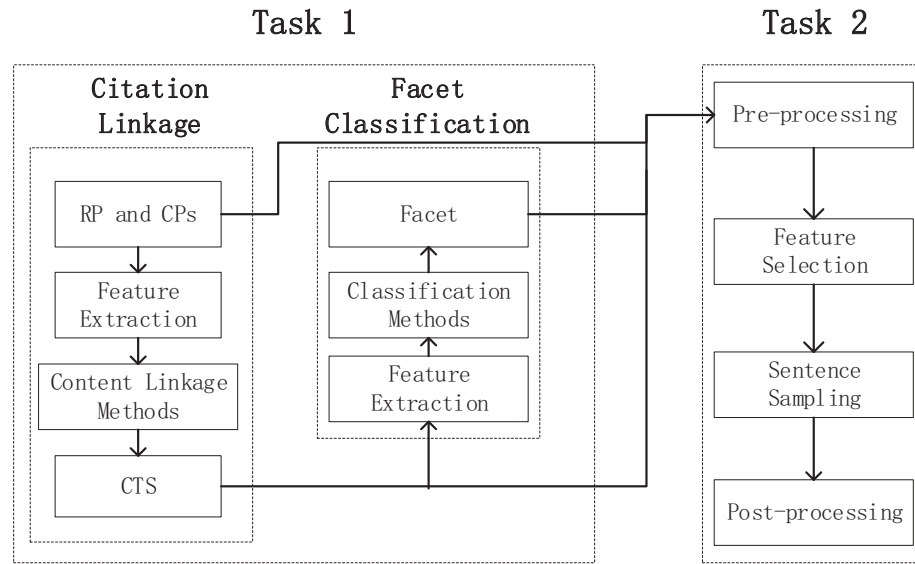


Fig. 1. Framework of Our System.

3.1 Word Mover's Distance

Word Mover's Distance (WMD) is a method for calculating the distance of two sentences or texts based on word vector and Earth Mover's Distance (EMD). WMD distance measures the dissimilarity between two textual documents as the minimum amount of distance that the embedded words of one document need to "travel" to reach the embedded words of another document [15]. We apply WMD as the measurement for similarity of two sentences and two texts in our system. Where, N and M are word number of two textual documents D and D' . w is word vector, and dim represents word vector dimension. d and d' are normalized bag-of-words vectors of D and D' .

$$\begin{array}{ccc}
 \left(\begin{array}{cccc} d_1 & w_{11} & \cdots & w_{1dim} \\ d_2 & w_{21} & \cdots & w_{2dim} \\ \vdots & \vdots & \ddots & \vdots \\ d_N & w_{N1} & \cdots & w_{Ndim} \end{array} \right) & & \left(\begin{array}{cccc} d'_1 & w_{11} & \cdots & w_{1dim} \\ d'_2 & w_{21} & \cdots & w_{2dim} \\ \vdots & \vdots & \ddots & \vdots \\ d'_M & w_{M1} & \cdots & w_{Mdim} \end{array} \right) \\
 \text{D} & & \text{D}'
 \end{array}$$

Fig. 2. Representation of two documents D and D'

After removing stop words, we first represent D and D' as two nBOW vectors d and d' . We then obtain word vector w of each word in D and D'. Finally we can obtain the representation of D and D' shown in Fig. 2. The goal of WMD is to incorporate the semantic similarity between individual word pairs (*e.g. President and Obama*), and the Euclidean distance of two words in the *word2vec* embedding space. The distance between word i and word j is $c(i, j) = \|w_i - w_j\|$. Word i and word j are from D and D' respectively. After getting d , d' and $c(i, j)$ we can use EMD algorithm to obtain the minimum WMD.

3.2 Task 1

Citation Linkage (Task 1A): The main processes are extracting features from RP and CPs, and using Content Linkage Methods to obtain CTS for each citance.

Feature Extraction: This is extracting features from RP and CPs, which contain Lexicons (high-frequency lexicon, LDA lexicon and co-occurrence lexicon), Sentence similarity (WMD similarity, IDF similarity and Jaccard similarity), Context similarities, Word vector, WordNet (jcn, lin, lch, res, wup and path similarity) and CNN (Convolutional Neural Network) similarity. We calculate the WordNet similarity between words in the two sentences to obtain a matrix. Then we select the maximum value in the matrix, and remove the corresponding row and column of the maximum value repeatedly until the matrix is null. Finally we add up all maximum values selected in each iteration to a sum value and the result is divided by $\sqrt{length_1 length_2}$ to be similarity between sentences. The process of computing Word vector similarity is the same as that of the WordNet similarity. CNN uses word vector as the input to obtain the probability of content linking from its output, and the output probability represents the similarity of input sentences [7]. Most features are used in our former work [7] except for Lexicon obtained by LDA model and the WMD applied for calculating Sentence similarities and Context Similarities.

In our previous work, we used LDA model only to train RP and CPs to obtain the LDA lexicon of 20 latent topics for files in each topic. We improve the LDA model to obtain better topic features. According to the LDA model

we denote a sentence S as an n -dimensional vector (LDA vector), such as $S = (x_1, \dots, x_i, \dots, x_n)$. x_i represents the probability of S which belongs to the i th topic. Every citance and CTS can be represented as n -dimensional vectors separately so that we could calculate their cosine similarity. We represent cosine similarity of LDA vector as LDA-cos. The larger cosine similarity is, the more similar they are. Compared with the old LDA method, the new LDA method not only considers the number of same words belonging to the same topic in citance and CTS, but also preserves the cohesion of topic distribution in them.

Besides, we use WMD to calculate the similarity of two texts for enriching similarity features.

Content Linkage Methods: We use two methods which are Voting Method and WMD Method. Voting Method means that final results are obtained by voting of all runs (which are the results given by features described in *Feature Extraction*). WMD Method means that the results come from the similarity calculated by WMD (we can call it WMD similarity). In the WMD similarity method, first we represent sentences as word vectors. Then we calculate the WMD similarity between citance and CTS using word vectors. WMD refers to the distance one specific sentence requires to transform to another, so the smaller the WMD is, the more similar the two sentences are.

Facet Classification (Task 1B): Our system mainly uses **Rule-based methods** and **Machine Learning Methods** based on multiple features for Task 1B. Rule-based methods contain Subtitle Rule (Sub), High Frequency Word Rule (HFW) and Subtitle and High Frequency Word Combining Rule (SubHFW). Rule-based methods construct rules based on features got from CTS, RP and CPs. As for Machine Learning methods, we apply SVM, Decision Trees (DT) and K-Nearest Neighbor (KNN) to obtain facet. Besides, we also train Random Forest (RF), Gradient Boosting (GB) and Voting methods to obtain facet, which are based on the idea of Ensemble Learning. The features used in machine learning methods contain *Location of Paragraph*, *Document Position Ratio*, *Paragraph Position Ratio* and *Number of Citations or References*. Finally we combine all the results to obtain a fusion result, which is called Fusion method.

3.3 Task 2

The main process for summary generation consists of **Pre-processing**, **Feature Selection**, **Sentence Sampling** and **Post-processing**.

Pre-processing: We need to correct some xml-coding errors firstly. Besides, we have to make some preparations such as document merging, sentence filtering and input file generation for hierarchical Latent Dirichlet Allocation (hLDA). We merge the content of RP and the citations into a document. And we will not extract the sentence in the abstract of RP except for that it is selected in Task 1A. Besides, all documents are converted to lowercase letters. We filter the corpus for removing some equations, figures, tables and so on. Then we generate input file for hLDA which contains word index and their corresponding frequencies.

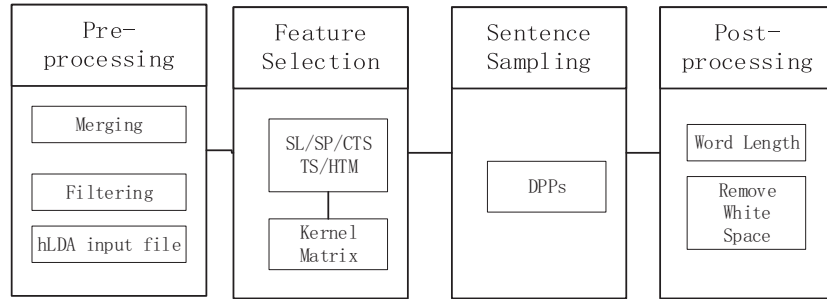


Fig. 3. Process of Task 2

Feature Selection: We choose Sentence Length (SL), Sentence Position (SP), CTS, Title similarity (TS) and Hierarchical Topic Model (HTM) as features in our system according to the work of Li L [3]. We use these features to calculate sentence quality. Besides, we use WMD similarity as sentence similarity, and combine it with sentence quality to construct kernel matrix of DPPs.

Sentence Sampling: We use DPPs to select sentences, which are elegant probabilistic models of global, negative correlations and mostly used in quantum physics to study the reflected Brownian motions. In our method, we only consider discrete DPPs and follow the definition of Kulesza A et al. [16]. We can enhance the diversity of summary by using DPPs. Furthermore, we also use Jaccard similarity to construct kernel matrix as a comparison for the effectiveness of DPPs based on WMD similarity.

Post-processing: We truncate the output summary to 250 words, and remove some white spaces in Post-processing.

4 Implementation and Experiments

We implement our system and use the official scripts to evaluate the training data using ten cross-validation in Task 1. Training-Set-2018 and Test-Set-2018 provided by official are training data and test data respectively in our system.

4.1 Task 1A

In our previous work, for syntactic information, we have three lexicons, two sentence similarities and two context similarities. All of them can measure sentence similarity [7]. For semantic information, we use word vector [7], WordNet and CNN. In this paper, we combine two feature representations (LDA vector and word vector) and two similarity calculation methods (EMD similarity and cosine similarity). We obtain two new methods: LDA-cos and WMD. We used the corpus crawled from "https://www.theguardian.com The Guardian" to train the word embeddings. The size of the corpus is 835 MB. As to experiments, we

choose 600 dimensions for LDA vector and 300 dimensions for word vector. The **Task 1A** methods are unsupervised. We have done some experiments under conditions of different numbers of sentences in the result. Then we choose the number used in our runs, which shows the best performance.

Besides, we also improve two feature fusion methods: Voting-1.0 and Jaccard-Focused in Li L et al. [7]. Except for some parameter changes, we add and delete some features of methods. Based on Voting-1.0 we obtain Voting-1.1, which replaces Jaccard context similarity with LDA-cos similarity. Based on Jaccard-Focused we obtain Jaccard-Focused-new, which adds jcn similarity and LDA-cos similarity. Table 1 shows the parameter settings of our methods.

Table 1. Parameter settings of Methods in Task 1A

Feature	V-1.1		V-1.0		V-2.0		J-F-new		J-F		J-C	
	W	P	W	P	W	P	W	P	W	P	W	P
IDF simmilarity	1	7	1	8	1	7	0.6	16	0.7	15	1.5	16
IDF context similarity	-	-	-	-	0.5	4	0.5	15	0.5	15	1	18
Jaccard similarity	1	7	1	12	1	3	JS	7	JS	7	-	-
Jaccard context similarity	-	-	1	12	0.5	8	0.7	16	0.7	15	1.5	15
word vector	1	6	1	10	0.5	8	0.5	26	0.5	25	-	-
Lexcion 2(LDA)	-	-	-	-	0.3	2	-	-	-	-	-	-
Lexcion 3(co-occurence)	-	-	-	-	0.4	2	0.2	23	0.2	25	0.5	15
jcn similarity	-	-	-	-	-	-	0.6	11	-	-	-	-
LDA-cos similarity	1	8	-	-	-	-	0.5	26	-	-	-	-
WMD similarity	-	-	-	-	-	-	-	-	-	-	-	-

In Table 1, W and P are Weight and Proportion respectively. V-1.1, V-1.0, V-2.0, J-F-new, J-F, J-C are Voting-1.1, Voting-1.0, Voting-2.0, Jaccard-Focused-new, Jaccard-Focused and Jaccard-Cascade methods repectively. JS means 10 fold of Jaccard Similarity. Owing to the performance of WMD similarity is very poor on the training data, WMD similarity is not adopted in our feature fusion methods.

Table 2. Performance of Methods in Task 1A

Method	Precision	Recall	F	Method	Precision	Recall	F
Voting-1.1	0.102	0.265	0.147	Jaccard-Focused-new	0.091	0.237	0.132
Voting-1.0	0.067	0.217	0.102	Jaccard-Focused	0.081	0.263	0.124
Voting-2.0	0.0838	0.271	0.128	Jaccard-Cascade	0.0.076	0.247	0.116

From Table 2, we find that the performance of Voting-1.1 method is better than Voting-1.0. It shows the validity of LDA-cos similarity. Besides, comparing to Jaccard-Focused method, the performance of Jaccard-Focused-new is much better.

4.2 Task 1B

Here, we mainly apply Rule-based Methods and Machine Learning Methods.

Rule-based Methods:

Subtitle Rule: We use the subtitles of CTS and citance to determine the facet. If the subtitles contain words of five predefined classes, we categorize CTS and citance as corresponding facet.

High Frequency Word Rule: We apply high frequency words obtained from five classes to classify CTS and citance. We first remove the common words, and then set a threshold for each facet.

Subtitle and High Frequency Word Combining Rule: We first apply Subtitle Rule to obtain the facet. If subtitles fail, we use High Frequency Words to obtain final facet.

Machine Learning Methods:

First we extract features from CTS and citance. The features are *Location of Paragraph*, *Document Position Ratio*, *Paragraph Position Ratio* and *Number of Citations or References* of CTS and citance, and they are put together in an 8-dimension vector. Second we train SVM, DT, KNN, RF, GB and Voting model with Training-Set-2016 and Training-Set-2017.

Table 3. Performance of Methods in Task 1B

Method	F_Score	Method	F_Score	Method	F_Score
Sub	0.716	GB	0.548	SVM	0.473
HFW	0.542	KNN	0.525	Voting	0.603
SubHFW	0.716	DT	0.462	RF	0.647

From Table 3, we can find that Sub, SubHFW, RF and Voting methods show better performance in our experiments. Owing to Sub Methods are highly related to subtitle, the method is full of uncertainty. In our submitted runs, we use RF, SubHFW, Voting and Fusion methods as our final methods for Task 1B.

Owing to the missing of some Citance.XML files in Test-Set-2018 released by the official, we cannot extract features of CTS. In this situation, we set a fixed initial value as features for Task 1B in submitted Test-Set-2018 runs.

4.3 Task 2

In this part, our system provides a sample method based on DPPs [7] to extract sentences when constructing a brief summary with no more than 250 words. Determinantal point processes (DPPs) are elegant probabilistic models of repulsion that origin in quantum physics and random matrix theory. The essential characteristic of a DPP is that these binary variables are negative correlated. As a result the sampling subset is a set of items that are diverse, this exactly encourages a number of techniques working with diverse sets, especially in the

information retrieval community . A summary generated by an automatic system requires the analogous principles: coverage of information, information significance, redundancy in information and cohesion in text. Thus, we associate these two objects together to build informative summaries through a sampling method based on DPPs by selecting diverse sentences from documents. It takes not only the ranking of the sentence quality themselves into account, but also the correlation between these sentences. This approach was once fully described before in [7] and was proven a competitive method based on the result feedback from the CLSciSumm-17.

As Task 2 requires a structured summary generated from the CTSs identified in Task 1A, we consider the CTS as one crucial feature described in section 3.3 to help select sentences. Besides, SP, SL, TS and HTM feature are also included. We try two specific metrics to measure the cohesion quantitatively: JACCARD calculates the proportion of same words precisely while WMD reflects the transition cost from one sentence to another. During our contrast experiment, we are looking forward to finding a best linear combination of qualities in order to capture more obvious characteristic for high-quality summary, and exploring relationship between sentences through comparison of different metrics for its redundancy.

The results below utilize Manual ROUGE values to evaluate our summaries. During the evaluation phase, CLSciSumm-18 has provided THREE kinds of criterion for option: the collection of citation sentences (the community summary), faceted summaries of the traditional self-summary (the abstract), and ones written by well-trained annotators (the human summary).

Take community summary for instance, we test SP (φ_0), SL (φ_1), TS (φ_2), HTM(φ_3) and CTS (φ_4) feature independently to figure out its own contribution at first. As the CTS feature (φ_4) is specifically designed, we tend not to present its individual performance, but record and observe the binary combination with every other basic feature.

Table 4. Binary Quality Combination Test

Run ID	φ_0	φ_1	φ_2	φ_3	φ_4	ROUGE1	ROUGE2
W-D-0	1	0	0	0	1	0.43652	0.23824
W-D-1	0	1	0	0	1	0.42104	0.19574
W-D-2	0	0	1	0	1	0.52800	0.37682
W-D-3	0	0	0	1	1	0.41193	0.18440

From Table 4, the best binary combination comes from TS (φ_2) and CTS (φ_4) features for WMD metric. One possible explanation is that the community summary itself has already included these citation sentences. With the title containing the essence of a paper, selected sentences following this ranking rule will definitely guarantee the overlapping on golden summaries.

Analogically, we conduct experiments on other two kinds of golden summaries, where the weights of parameters appear slightly different. Tables 5-7,

Table 5. Performance On Community Summary

Run ID	φ_0	φ_1	φ_2	φ_3	φ_4	ROUGE1	ROUGE2
W-D-0	0	0	1	0	1	0.52800	0.37682
W-D-1	0	0	0	1	1	0.41193	0.18440
W-D-2	1	0	2	0	2	0.44158	0.25259
W-D-3	1	1	2	0	2	0.43992	0.24908
J-D-0	0	0	1	0	1	0.52552	0.37333
J-D-1	0	0	0	1	1	0.41283	0.18438
J-D-2	1	0	2	0	2	0.44104	0.24653
J-D-3	1	1	2	0	2	0.42219	0.22141

Table 6. Performance On Self-summary

Run ID	φ_0	φ_1	φ_2	φ_3	φ_4	ROUGE1	ROUGE2
W-D-0	1	0	0	0	0	0.39625	0.18644
W-D-1	0	1	0	0	0	0.36460	0.14273
W-D-2	0	0	1	0	0	0.38662	0.17437
W-D-3	0	0	0	1	0	0.30555	0.07241
J-D-0	1	0	0	0	0	0.39630	0.19019
J-D-1	0	1	0	0	0	0.35020	0.11507
J-D-2	0	0	1	0	0	0.38434	0.17296
J-D-3	0	0	0	1	0	0.30237	0.08561

Table 7. Performance On Human Summary

Run ID	φ_0	φ_1	φ_2	φ_3	φ_4	ROUGE1	ROUGE2
W-D-0	0	0	1	0	1	0.41884	0.19276
W-D-1	0	0	0	1	1	0.34337	0.09636
W-D-2	1	0	2	0	2	0.40278	0.16916
W-D-3	1	1	2	0	2	0.41598	0.18429
J-D-0	0	0	1	0	1	0.41900	0.19167
J-D-1	1	0	0	0	1	0.39866	0.15321
J-D-2	2	0	3	0	3	0.43504	0.25430
J-D-3	2	1	3	0	3	0.42219	0.22141

present the weights and results of the three golden summaries: the best binary combinations go to the same tendency. However, when it comes to human summary, the more parameters are involved, the higher ROUGE F-score it reaches. Unfortunately, for community summary, when we desire a further exploration on binary combination, any additional attribute performs adversely. The phenomena of same best combination may be interpreted implicitly that no matter whether the sentences are cited otherwise or the summaries are written by annotators, the two both are from the perspective of readers. *There are a thousand Hamlets in a thousand people's eyes.* As for the self-summary (the abstract), every binary combination with CTS (φ_4) feature are not that satisfied, so we present each individual contribution of other statistical or topic features. Perhaps although we have tried our best to follow the writers, there always exists a narrow gap between our readers' comprehension and writers' original intention. In general, despite the two diversity metrics are somehow evenly matched on this dataset, the best result in Table 5, the 1th row comes from WMD metric, thus we firmly believe the newly proposed algorithm is just on its way, still remains full potential to be discovered.

5 Conclusion and Future Work

In this paper, we propose some new methods to improve the performance of Task 1 and Task 2 based on our former work, especially in similarity calculation. We apply WMD method and LDA-cos to calculate similarity and generate summaries. In future, we will continue to improve these methods and incorporate new methods based on the official results by CLSciSumm-18.

Acknowledgements

This work was supported by National Social Science Foundation of China [grant number 16ZDA055]; National Natural Science Foundation of China [grant numbers 91546121, 71231002]; EU FP7 IRSES MobileCloud Project [grant number 612212]; the 111 Project of China [grant number B08004]; Engineering Research Center of Information Networks, Ministry of Education; Beijing BUPT Information Networks Industry Institute Company Limited; the project of Beijing Institute of Science and Technology Information; the project of CapInfo Company Limited.

References

1. CL-SciSumm 2018 Homepage, <http://wing.comp.nus.edu.sg/birndl-sigir2018/>.
2. Chandrasekaran M K, Jaidka K, Mayr P. Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017)[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2017: 1421-1422.

3. Li L, Zhang Y, Chi J, et al. UIDS: A Multilingual Document Summarization Framework Based on Summary Diversity and Hierarchical Topics[M]//Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer, Cham, 2017: 343-354.
4. Jaidka K, Chandrasekaran M K, Rustagi S, et al. Insights from CL-SciSumm 2016: the faceted scientific document summarization Shared Task[J]. International Journal on Digital Libraries, 2017: 1-9.
5. Jaidka K, Chandrasekaran M K, Jain D, et al. The CL-SciSumm shared task 2017: results and key insights[C]//Proceedings of the Computational Linguistics Scientific Summarization Shared Task (CL-SciSumm 2017), organized as a part of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017). 2017.
6. Ma S, Xu J, Wang J, et al. NJUST@ CLSciSumm-17[C]//Proc. of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2017). Tokyo, Japan (August 2017).
7. Li L, Zhang Y, Mao L, et al. CIST@ CLSciSumm-17: Multiple Features Based Citation Linkage, Classification and Summarization[C]//Proc. of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2017). Tokyo, Japan (August 2017).
8. Zhang D, Li S. PKU@ CLSciSumm-17: Citation Contextualization[C]//Proc. of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2017). Tokyo, Japan (August 2017).
9. Pramanick, Aniket, et al. "SciSumm 2017: Employing Word Vectors for Identifying, Classifying and Summarizing Scientific Documents."
10. Aburaed, Ahmed, et al. "LaSTUS/TALN@ CLSciSumm-17: cross-document sentence matching and scientific text summarization systems." (2017).
11. Lauscher, Anne, Goran Glava, and Kai Eckert. "University of Mannheim@ CLSciSumm-17: Citation-Based Summarization of Scientific Articles Using Semantic Textual Similarity." (2017): tba.
12. Felber, Thomas, and Roman Kern. "Graz University of Technology at CL-SciSumm 2017: Query Generation Strategies."
13. Karimi, Samaneh, et al. "University of Houston@ CL-SciSumm 2017: Positional language Models, Structural Correspondence Learning and Textual Entailment."
14. Prasad, Animesh. "WING-NUS at CL-SciSumm 2017: Learning from syntactic and semantic similarity for citation contextualization." Proc. of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2017). Tokyo, Japan (August 2017). 2017.
15. Kusner M, Sun Y, Kolkin N, et al. From word embeddings to document distances[C]//International Conference on Machine Learning. 2015: 957-966.
16. Kulesza A, Taskar B. Determinantal point processes for machine learning[J]. Foundations and Trends in Machine Learning, 2012, 5(23): 123-286.