2016 12th International Conference on Semantics, Knowledge and Grids

Summarizing Related Work through Citations

Jingqiang Chen, Hai Zhuge

Nanjing University of Posts and Telecommunications, 210023, Nanjing, China

Key Lab of Intelligent Information Processing, Institute of Computing Technology, University of Chinese Academy of Sciences, 100190, Beijing, China

Systems Analytics Research Institute, Aston University, Birmingham, UK.

[1]cjq@njupt.edu.cn [2]zhuge@ict.ac.cn

Abstract—**Automatically generating the related work section for a writing paper is useful for researchers as it can save a lot of time and avoid missing related works. The related work section of scientific paper usually introduces other researchers' work published before and makes comparisons with the current author's work. This paper proposes an approach to automatically generating related work by comparing the main text of the writing paper with the citations to the references from other papers. Our approach firstly collects the papers that cite the reference papers of the writing paper and extracts the corresponding citation sentences to form a citation document. It then makes a summarization that makes a comparison between the citation document and the writing paper's abstract, introduction and conclusion. It extracts the representative keywords from the citation document and the writing paper, and constructs a graph of the keywords. The discriminated nodes in the graph are figured out, and then the minimum Steiner tree that covers the nodes is extracted. A summary is generated by extracting the sentences covering the Steiner tree. The experiments show that our approach outperforms three baselines MEAD, ReWoS and ARWG according to ROUGE evaluation results.**

I. Introduction

Related work has some patterns that facilitate representation and understanding [1]. Let's look at an example taken from reference [2]: The title of the paper is "Using Citations to Generate Surveys of Scientific Paradigms". Its related work section firstly introduces relevant work from multiple aspects, including analysis of citation networks, citation categorization, and etc. At the end of the section the author introduces the difference and the main contribution of the author's work. The title of the paper is used as the topic of the related work section. Each paragraph groups the references and introduces the topic from different aspects. The difference and contribution of the author's work are emphasized in an independent paragraph. Other similar examples of related work sections are in references [3] and [4].

According to these examples, the related work section consists of three parts: 1) comments on the references of the writing paper; 2) introduction of the difference; and 3) the contribution of the writing paper.

Comments on the references of a paper can be found in other papers that also cite the references, or in the main text of the reference papers. For example, a similar citation of "Nanba and Okumura (1999) discuss citation categorization to support a system for writing a survey." can be found in reference [5]. These comments can be used to form the related work section.

> Previous work has focused on the analysis of citation and collaboration networks (Teufel et al., 2006; Newman, 2001) and scientific article summarization (Teufel and Moens, 2002). Bradshaw (2003) used citation texts to determine the content ……
>
> ……..
>
> Nanba and Okumura (1999) discuss citation categorization to support a system for writing a survey. Nanba et al. (2004a) automatically categorize citation sentences into three groups using pre-defined phrase-based rules. …….
>
> ………
>
> Our aim is not only to determine the utility of citation texts for survey creation, but also to examine the quality distinctions between this form of input and others such as abstracts and full text ……

Fig. 1. A part of the related work section of the paper "Using Citations to Generate Surveys of Scientific Paradigms".

Different from the main text of the reference papers, the citations that cite a reference paper are concise and general summarizations. Sometimes, the citations contain information that does not appear in the main text of a paper [6]. Some researchers use citations to generate a summary of a paper [5] or a survey of a domain [2].

The difference and the contribution of a paper can be found in the main text of the paper. For example, a similar sentence of the sentence "Our aim is not only to determine the utility of citation texts for survey creation, but also to examine the quality distinctions" in the related work section of reference [2] also appears in the abstract of the paper.

This paper automatically creates the related work section by extracting the appropriate sentences from the main text of the writing paper and the citation sentences that cite the reference papers. The created related work section serves as a draft for the author. A comparative style of summary is used to generate the summary.

Comparative summarization aims to get not only common points of two documents, but also the differences of the two documents. This kind of summarization is fairly suitable for related work section creation since the related work section discusses different contributions of the reference papers and

978-1-5090-4795-6/16 $31.00 © 2016 IEEE
DOI 10.1109/SKG.2016.31

highlights the difference and the contribution of the writing paper.

Our summarization approach goes as follows: A feature graph is built to represent the citation document and the main text document. The citation document consists of the citation sentences. The main text document consists of the abstract, introduction and conclusion section of the writing paper. In the graph, the nodes represent keywords and the edges represent association between keywords. The common keywords occurred in both documents represent the commonality and connect the graph. The discriminative keywords represent the differences of the two documents. A minimum Steiner tree that covers the discriminative keywords and the common keywords is extracted. Finally, the sentences that cover the Steiner tree are extracted to form a summary. The Steiner tree incorporates the commonality and the difference of the citation document and the main text document. The resultant summary can get the minimum sentences that cover the most information in the texts.

For evaluation, the author-written related work section is compared with the summary created by the automatic approaches. *ROUGE* evaluation method is used. The results show that our approach outperforms four baselines *MEAD*, *ReWoS*, *LexRank*, and *ARWG*.

II. Related Work

Several works for generating related work section mainly extract text from the main text of the references. A related work summarization system named *ReWoS* takes in a set of keywords arranged in a hierarchical structure and extract important sentences from the references to create a summary [7]. An optimization approach was used to automatically create related work section [8]. They use a PLSA model to split the sentence set into different topic-biased parts, and then learn the importance of the sentences which is used to generate the related work section. Widyantoro and Amin's work aims to identify and classify citation sentence from scientific articles for related work summarization [9].

Some researchers use citations to create the summary of a paper and to create the survey of a research domain. Elkiss etc find that citations have different information not contained in the abstracts [6]. Based on the finding, Qazvinian and Radev employ the citations to create the summary for the scientific paper [3][5][10][11]. Moreover, Mohammad et al. use citations to generate a survey of a scientific domain [2]. These works show that citations have special information and can be used to create summaries.

Comparative summarization has been studied for several years [12-18]. One strategy of comparative summarization only considers the notable difference of the two documents but without considering their commonality. Wang et al. model the summarization as an optimization problem that tries to minimize the conditional entropy of the sentence membership given the selected sentence set. Another strategy considered both commonalities and differences of documents when selecting representative sentences [12]. Typically, two documents are related to each other, i.e., they share some common aspects; nevertheless, their focus on these aspects might be different. Based on this observation, several methods have been reported to generate comparative summaries. One representative work [18] considers semantic-related cross-topic concept pairs as comparative evidences, and topic-related concepts as representative evidences. Comparative summarization is also applied to Patent domain to discriminate two patents [19].

Citation is a kind of semantic link that summarizes a piece of text from author's point of view. Summarization was regarded as a process of general citation and a method for general summarization based on the connotation and extension through citation was proposed [1]. The systematic theory on semantic link network was established to study the semantic characteristics of the self-organized system [20][21][22][23][24][25][26].

This paper uses citations to do comparative summarization to generate related work section. The main contribution of the paper is as follows:

1) We propose to use citations that cite the references papers in other paper to create the related work section. The citations of the references are extracted and are then compared with the main text of the paper.

2) We propose a graph-based comparative summarization approach where the commonality of two documents is represented as common nodes in the graph and the differences of two documents are represented as the discriminated nodes. Differences and commonalty of the two documents are identified and the corresponding sentences are extracted to form a summary.

III. The Summarization Approach

Fig. 2 shows the architecture and the pipeline of our summarization system named *RWS-Cit* which consists of three main modules: citation collecting module, data preprocessing module, and summarization module. The summarization module is the core module of *RWS-Cit*. The other two modules are to prepare data for the summarization module.

The *RWS-Cit* system aims to automatically generate the related work section of a paper, assuming that the title of the paper, the abstract of the paper, the introduction of the paper, the conclusion of the paper, and the reference list of the paper have been given the author. We only use the abstract, the introduction and the conclusion section for that these parts contain dense and key information of the writing paper. The keywords of the title of the writing paper are used as the topics of the related work section. The automatically generated related work section serves as a draft for the author to write his/her own related work section.

The citation sentences distribute among the papers that cite the reference papers, and those citing papers spread across the web. Given a list of paper titles and paper authors, the citation collecting module automatically collects the citing papers from the web, and then extracts citation sentences from the citing papers.

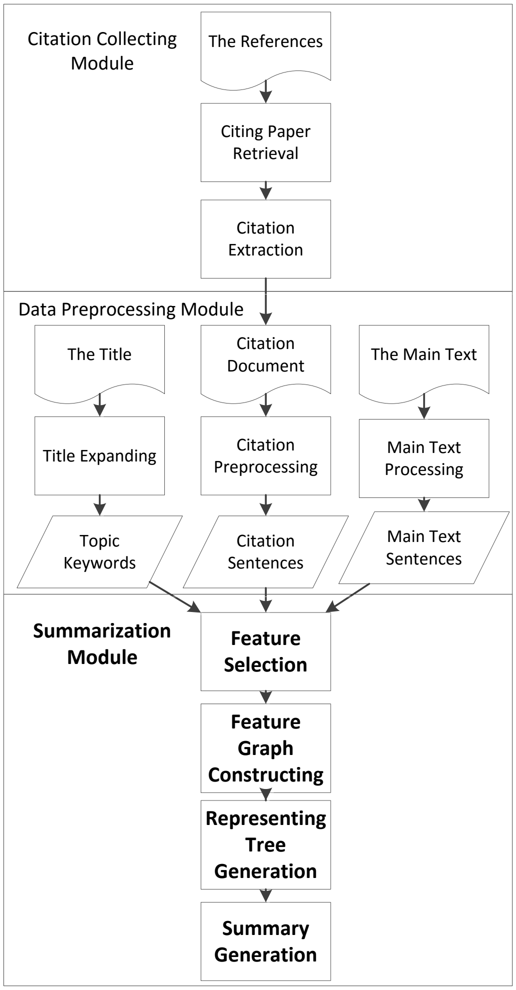

Fig. 2. The architecture of the summarization system *RWS-Cit*

The data preprocessing module preprocesses the titles, the main texts of the writing paper, and the citation sentences. The keywords of the title of the writing paper are extracted and expanded. The main text document is splitted to sentences which are then parsed.

The summarization module employs a comparative summarization method to automatically create the related work section from the preprocessed data. The module consists of four submodules: 1) Feature section; 2) Feature graph constructing; 3) Steiner tree generating; and 4) summarization. This module is the core module of the system

IV. The Citation Collecting Module

Citation Collecting Module collects citation sentences from the web. The model inputs a list of reference papers including the titles, the authors and the publishing years. The output of the model is a set of citation sentences that cite the reference papers. The module consists of two steps: 1) the citing papers retrieval step; and 2) the citation sentences extracting step.

The citing papers retrieval step retrieves the citing papers from scholar searching engines for each paper in the reference list provided by the author.. Google Scholar (http://scholar.google.com) and Baidu Scholar (http://xueshu.baidu.com) are two scholar searching engines under consideration. We choose Baidu Scholar for that it is easier for us to access. Baidu Scholar provides an open interface for scientific retrieval. Given the title of a paper as the query, it returns a web page about a list of citing papers that cite the paper in html format. We download the html-format web pages which are then parsed. The downloading links of the papers are extracted. And the papers are downloaded directly through the links.

The downloaded papers are in pdf format. To extract citation sentences, the papers are transformed into txt format through the pdfbox library (downloaded at http://pdfbox.apache.org/), an open-source library written in Java to transform a document from pdf format to txt format. The lines of each transformed txt-format paper are concated to a text block. The text blocks are splitted to sentences with *StanfordCoreNLP* toolkit [27].

Citation sentences are extracted from the preprocessed texts. Accurate citation extraction is an extremely challenging task due to variations in the use of field separators. Most of current works on citation extraction are knowledge-based [28][29], rule-based [30][31], and machine-learning-based [32]. The rule-based citation extraction approach is employed in *RWS-Cit* system. The rule-based approach defines rules or templates in advance, and applies the rules or templates to extract the citations. In our case, the cited papers of the citations we are to extract are known, so it is easier to define rules. And the precision is more important than the recall in our case. We want that the extracted citation sentences are the right sentences that cite the right paper. The recall can be compensated by the number of papers we extract citations from. For each reference paper in the input, *R* (usually set to 5) citation sentences are retrieved from the web. The collecting model exceeds from each reference paper in the input until *R* citation sentences are extracted.

For a writing paper, the collected citations that cite the references of the paper are put together to form a citation document.

V. The Data Preprocessing Module

The inputs of the model are the title of the target paper, the citation document and the main text document. The citation document consists of the citations. The main text document consists of the abstract section, the introduction section and the conclusion section of the writing paper.

A. Title Preprocessing

This module preprocesses to extract topics from the title of the paper. Different paragraphs of the related work section usually talk about different topics. And the topics can be got from the title of the paper. For example, the related work section exemplified in Section 1 talks about "citations", "surveys", and "scientific paradigms" which can be deduced from the title.

Nouns or noun phrases are the keywords of the title and contain key information. We first extract nouns and noun phrases from the title. Each noun or noun phrase is treated as a topic keyword.

The topic keywords are then expanded by retrieving synonyms from a thesaurus. Researchers may use different words to refer to a same thing. For example, the word "survey" is the synonym of the word "review". A thesaurus needs to be constructed. There are two approaches to build the thesaurus: 1) the knowledge-based approach; and 2) the corpus-based approach [33]. The knowledge-based approach uses the knowledge based resources such as WordNet to get synonym from. This is an open-domain approach. The corpus-based approach uses a domain-specific corpus and co-occurrence-based approaches (such as PMI, Cosine) to discover similarity between words. We choose the corpus-based approach for that our experiments focus on the Computational Linguistics domain.

We uses the abstracts and titles of a collection of scientific articles in a Computational Linguistics domain as the corpus, puts the abstract and title of a scientific paper together and treats them as an item set. We collect about 20000 papers of pdf format in ACL Anthology Network (http://www.aclanthology.net). The Cosine metric is employed to compute the relevancy between two keywords. A keyword co-occurrence base is constructed [34] [35].

B. *Citations and Main Text Preprocessing*

Citation sentences and the main text of the papers are splitted to sentences, and nouns in the sentences are extracted. The NLP toolkit used is StanfordCoreNLP.

Citation sentences are put together to form a citation document. The main text of the writing paper are put together to form a main text document. In the system, we default use the abstract section, the introduction section and the conclusion section to extract summary.

VI. The Summarization Module

A. *Different Feature Selection*

We refer features to keywords that can represent a document or a sentence. And nouns are used as keywords in our system.

Discriminative Features denote differences between the citation document and the main text document, and refer to keywords that can discriminate the two documents.

Formally, suppose we have *t* feature variables from the two documents, denoted by $\{x_i|x_i \in F\}$, where F is the full feature index set, having |F| = t. We have the document variable, $D = \{d_1, d_2\}$. The problem of feature selection is to select a subset of features to accurately predict the document variable *D*. There are various ways to perform feature selection, e.g., information theory based methods (such as point-wise mutual information), and statistical methods (such as $\chi 2$ statistics and Cosine statistics). Our approach uses mutual information as the feature selection method as it has been successfully applied to the field of text processing [36]. Equation (1) is the equation to compute PMI between the document d_i and the feature *f*. Since there are only two documents, $P(d_i)$ is computed by the word count in d_i divided by the total word count in the two documents. *P*(*f*) is the count of feature f in two documents divided by the total word count.

Equation (2) and equation (3) compute the average PMI score and the maximum PMI score of the feature *f* among two documents.

$$PMI(f, d_i) = \log \frac{P(\mathrm{f} \wedge \mathrm{d_i})}{P(f) \times P(d_i)} \quad (1)$$

$$PMI_{avg}(\mathrm{f}) = \sum_{i=1}^{m} \mathrm{P}(\mathrm{d}_i) PMI(\mathrm{f}, \mathrm{d}_i) \quad (2)$$

$$PMI_{\max}(f) = \max_{i=1}^{m} \{PMI(f, d_{\mathrm{i}})\} \quad (3)$$

The features of highest PMI values are selected out as different features. Different features are used to represent the differences of the documents.

B. *Feature Graph Constructing*

A feature graph is constructed to represent the two documents. Nodes in the graph represent the features, and edges in the graph represent the relationship between features. Common nodes represent the topic keywords that occur in both documents. The different nodes represent the different features..

The following equation is used to compute the weight of relevancy between two nodes v_1 and v_2.

$$EdgeScore_D(v_1, v_2) = 2 \times \frac{\left|\{(\mathrm{v}_1, \mathrm{v}_2) \mid \mathrm{v}_1 \in \mathrm{D}, \mathrm{v}_2 \in \mathrm{D}\}\right|}{\left|\{v_1 \mid v_1 \in D\}\right| \times \left|\{v_2 \mid v_2 \in D\}\right|} \quad (4)$$

In Equation (4), $EdgeScore(v_1,v_2)$ denotes the relation score of the edge between v_1 and v_2. $|\{v_1|v_1 \in A\}|$ and $|\{v_2|v_2 \in D\}|$ denote the frequencies of v_1 and v_2 in document *D*, respectively. $|\{(v_1, v_2)|v_1 \in D, v_2 \in D\}|$ represents the number of times that v_1 and v_2 appear in the same sentence of *D*. Given two documents D_1 and D_2, we connect v_1 and v_2 if their averaged linkage score on both documents exceeds a predefined threshold τ.

C. *Representing Tree Generation*

The different features obtained from the feature selection submodule represent the difference of the two documents. The different features may not be well connected in the feature graph, so we have to discover the relationship among the different features. This can possibly be achieved by connecting the different nodes and the nodes shared by the two documents. For presentation purpose, the generated summary should include the minimum number of features and convey the major commonalities/differences.

We formulate it as the minimum Steiner tree problem to connect the features in a meaningful way. Given a graph *G* and a subset of vertices *S*, a Steiner tree of *G* is defined as the subtree of *G* that contains *S* with the minimum number of edges.

The problem of Minimum Steiner Tree is an NP-hard problem [37]. A reasonable approximation can be achieved by finding the shortest path from the root node to the terminal nodes and then combining the paths. To find the root node is the key step. Once the root node is found, the problem is simplified to a shortest path finding problem. We use the topic keywords extracted from the titles and expanded in Preprocessing Module as the root nodes. The citation document and the main text document share the topic keywords.

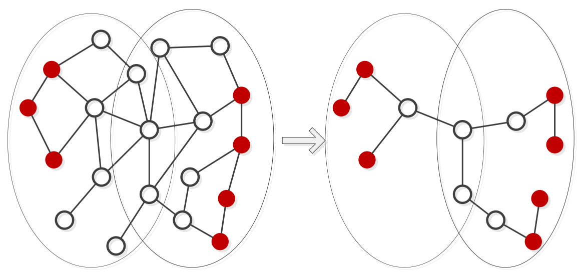

Fig. 3. The representing graph and the generated Steiner Tree

We use Floyd algorithm to solve the problem. Once the Steiner tree is generated, we can get a concise feature-based comparative summary of given scientific documents.

D. *Summarization*

The Steiner tree provides the basis for generating comparative summaries of the citation document and the main text document. We aim to select the minimum set of sentences from the original documents. The features in the Steiner tree can be fully covered by the sentences. Each sentence can be represented as a subgraph of the entire graph. The Steiner tree is regarded as a subgraph. Therefore, the problem can be formulated as selecting the minimum set of subgraphs that cover the Steiner tree.

The set cover problem is also an NP-hard problem. We propose a greedy algorithm according to the following rule: at each time, choose the sentence that contains the largest number of uncovered elements.

VII. *EVALUATIONS*

A. *Building the Experiment Dataset*

We refer features to keywords that can represent a document or a sentence

Since there is no benchmark or existing dataset for evaluating the generation of the related work section, we manually collect three papers, i.e., reference [2], reference [3] and reference [4] as experiment papers and build the experiment dataset as follows.

1) Building the citation document for each experiment paper. *RWS-Cit* collects citation sentences from Web automatically given the references of an experiment paper. For comparison, we also manually build the citation document. For each reference paper in the experiment paper, we collect the papers that cite the reference paper from Baidu Scholar. We download the top 5~10 papers from the result list and extract the corresponding citation sentences. For either experiment papers, we collect about 30 citation sentences and form a citation document. It is a time consuming task. We ask three undergraduate students to do this.

2) Building the main text document for each experiment paper. We extract the abstract section, the introduction section and the conclusion section as the main text of the papers for experiments.

3) Building the gold standard data. For each experiment paper, the existing related work section is extracted and is used as the gold standard summaries.

4) Preprocessing the data. For each document, we use StanfordCoreNLP [27] to do sentence splitting, POS-tagging and parsing. The nouns are kept as keywords, and then the preprocessing module of *RWS-Cit* is applied.

For evaluation, the *ROUGE* is used as the evaluation metric. *ROUGE* has been traditionally used to evaluate the automatically created summaries based on the n-gram overlap between the automatically created summaries and the gold standard summaries [38]. Two versions of the *ROUGE* metric, *ROUGE*-1 and *ROUGE*-2, are employedin our experiments.

B. *Comparing with existing methods*

One general-purposed baseline method and two domain-specific base-line methods are used for comparison.

1) The first one is the widely used and freely available state-of-art *MEAD* system [39]. MEAD is a general-purposed text summarization system which can be applied to any type of text applications other than scientific documents. *MEAD* implements a battery of summarization algorithms, including centroid-based and query-based methods. We use the default setting of *MEAD*. the sentences are ranked by length, position and centroid in the default setting of *MEAD*. To apply *MEAD* to related work section generation, we use the main text of the reference papers and the titles of the experiment paper as input.

2) The second one is *ReWoS*, which extracts sentences from the main texts of the references to generate a related work section. The method assumes the related work section has a hierarchical structure. It manually builds a hierarchical structure of topic in advance, based on which the sentences are extracted from the main text of the reference papers. We implement the approach by build two-layer topic hierarchy from the title of the writing paper.

3) The third one is *ARWG*, which first exploits a PLSA model to split the sentence set of the given papers into different topic-biased parts, and then applies regression models to learn the importance of the sentences. At last ARWG employs an optimization framework to generate the related work section.

The three baseline methods extract sentences from the main text of reference papers to create summaries. In contrast, our method employs citations to generate the related work section.

Table 1 shows the *ROUGE* scores of *RWS-Cit*, **RWS-Cit*, *MEAD*, *ReWoS*, and *ARWG*. **RWS-Cit* is our approach using the manually collected citations to do summarization. According to the table, *RWS-Cit* and **RWS-Cit* perform better than the three baseline methods. What surprises us is that the scores of *RWS-Cit* are not far from the scores of **RWS-Cit*. And some scores of *RWS-Cit* are higher than that of **RWS-Cit*.

Table 1. The ROUGE scores of RWS-Cit, *RWS-Cit, MEAD, ReWoS, and ARWG.

Metric	*RWS-Cit*	**RWS-Cit*	*MEAD*	*ReWoS*	*ARWG*
ROUGE-1 F-measure	0.45281	0.50151	0.33693	0.37005	0.37511
ROUGE-1 Recall	0.55934	0.52755	0.30567	0.36798	0.37420
ROUGE-1 Precision	0.39374	0.48451	0.40315	0.40542	0.41228
ROUGE-2 F-measure	0.18008	0.16592	0.06848	0.08045	0.09802
ROUGE-2 Recall	0.21846	0.17437	0.05891	0.08108	0.07210
ROUGE-2 Precision	0.15788	0.16055	0.08648	0.08735	0.10205

Nanba and Okumura (1999) discuss citation categorization to support a system for writing a survey. Nanba et al. (2004a) automatically categorize citation sentences into three groups using pre-defined phrase-based rules. Based on this categorization a survey generation tool is introduced in Nanba et al. (2004b). They report that co-citation (where both papers are cited by many other papers) implies similarity by showing that the textual similarity of co-cited papers is proportional to the proximity of their citations in the citing article.

Elkiss et al. (2008b) conducted several experiments on a set of 2,497 articles from the free PubMed Central (PMC) repository.[1] Results from this experiment confirmed that the cohesion of a citation text of an article is consistently higher than the that of its abstract. They also concluded that citation texts contain additional information are more focused than abstracts.

Nakov et al. (2004) use sentences surrounding citations to create training and testing data for semantic analysis, synonym set creation, database curation, document summarization, and information retrieval. Kan et al. (2002) use annotated bibliographies to cover certain aspects of summarization and suggest using metadata and critical document features as well as the prominent content-based features to summarize documents. Kupiec et al. (1995) use a statistical method and show how extracts can be used to create summaries but use no annotated metadata in summarization.

Our aim is not only to determine the utility of citation texts for survey creation, but also to examine the quality distinctions between this form of input and others such as abstracts and full texts—comparing the results to human-generated surveys using both automatic and nugget-based pyramid evaluation (Lin and Demner-Fushman, 2006; Nenkova and Passonneau, 2004; Lin, 2004).

Fig.4. The related work section of the paper “Using Citations to Generate Surveys of Scientific Paradigms”

[1] Previous work has studied and used citation sentences in various applications such as: scientific paper summarization (elkiss et al., 2008;), automatic survey generation (nanba et al., 2000;), citation function classification (nanba et al., 2000;teufel et al., 2006; siddharthan and teufel, 2007;teufel, 2007).

[2] They must learn about a new discipline “on the fly” in order to relate their own expertise to the proposal. our goal is to effectively serve these needs by combining two currently available technologies: 1) Bibliometric lexical link mining that exploits the structure of citations and relations among citations; and 2) summarization techniques that exploit the content of the material in both the citing and cited papers.

[3] Qazvinian and radev (2008) argue that citation texts are useful in creating a summary of the important contributions of a research paper .the citation text of a target paper is the set of sentences in other technical papers that explicitly refer to it (elkiss et al., 2008a).

[4] However, there is no previous work that uses the text of the citations to produce a multi-document survey of scientific articles. furthermore, there is no study contrasting the quality of surveys generated from citation summaries-both automatically and manually produced-to surveys generated from other forms of input.

[5] We generated surveys of a set of Question Answering (QA) and Dependency Parsing (DP) papers, their abstracts, and their citation texts using four state-of-the-art summarization systems (C- LexRank, C-RR, LexRank, and Trimmer).

[6] The notion of text level categories or zoning of technical papers-related to the survey components enumerated above-has been investigated previously in the work of nanba and kan (2004b) and teufel (2002).

[7] The area of that curve is the probability that a randomly drawn positive example has a higher decision function value than a random negative example. we define a baseline method based on the method proposed by nanba and okumura (nanba and okumura 1999).

[8] nanba et al. (2004a) automatically categorize citation sentences into three groups using pre-defined phrase-based rules.

[9] Even though prior work (teufel etal., 2006) argues that citation text is unsuitable for summarization, we show that in the framework of multi-document survey creation, citation texts can play a crucial role.1.

[10] The first stream of research applies a rule-based strategy based on predefined cue-words or phrases set in a decision tree classification to classify extracted citations (garzone, 1997; nanba, kando, &okumura, 2000).

Fig.5. The summarization results of RWS-Cit for the paper titled “Using Citations to Generate Surveys of Scientific Paradigms”

Fig. 4 shows the original related work section of one of the experiment papers. The first three paragraphs talk about the work of its reference papers, and the last paragraph emphasizes the contribution of the experiment paper.

Fig. 5 shows the automatic created related work section of *RWS-Cit* for the same paper. The summary is formed by seven citation sentences from the citation document and two sentences from the main text of the experiment paper. Which sentences to be selected is determined by the *RWS-Cit* method.

Discriminate features in the feature graph ensure that for either document at least one sentence can be selected.

The discussion of reference papers in Fig. 3 and Fig.4 are highly overlapped in some important information. For example, the sentence "Nanba and Okumura (1999) discuss citation categorization to support a system for writing a survey." in Fig. 4 is the same as the sentence "In other work, (Nanba et al., 2004b; Nanba et al., 2004a) analyze citation sentences and automatically categorize citations into three groups using 160 pre-defined phrase-based rules." in Fig. 5. The reason is that the comments of the reference papers in other papers are probably the same as a new writing paper under one topic. While other three baseline methods extract text from the main text of the reference papers, not from the citations.

The discussion of contribution part in Fig. 4 and Fig. 5 are also highly overlapped. For example, the sentence "Our aim is not only to determine the utility of citation texts for survey creation, but also to examine the quality distinctions between this form of input and others such as abstracts and full texts" is the same as the last two sentences in Fig. 5, extracted from the abstract of the experiment paper. In contrast, the summaries created by the three baseline methods do not contain the discussion part of the new writing paper.

*RWS-Cit uses manually collected citations to generate summaries, while RWS-Cit uses automatically collected citations to generate summaries. The quality of the manually collected citations is better than that of the automatically collected citations, due to limits of NLP techniques. However, it shows that the difference in the quality of citations does not influence the scores of the summaries significantly. This is because different citation sentences about a paper possibly talk about same things. Once a citation is extracted, most information is covered by the citation. Both *RWS-Cit and RWS-Cit outperform the three baseline methods.

In summary, *RWS-Cit* employs the comparative summarization approach to extracting the differences and commonalities from the citation documents and the main text of the new writing paper. The comments of the reference papers and the discussion of the contribution of the writing paper can be extracted spontaneously. The other three baseline methods extract text from the main text of the references and do not consider the differences and the commonalities. This is the main reason that *RWS-Cit* outperforms *MEAD*, *ReWoS*, and *ARWG*.

C. *Evaluation with different sections*

RWS-Cit uses citations of reference papers in other papers and the main text of the new writing paper to generate the related work section. In the experiment, we use the main text of the reference papers and the abstract plus the conclusion section of the new writing paper to create summaries, to see whether the main texts of reference papers and other sections of the new writing paper can be used and how it performs. The results are shown in Table 2.

In Table 2, *RWS-Cit-AIC* uses the citation documents and the abstract, the introduction and the conclusion section of the new writing paper. *RWS-Cit-FULL* uses the abstract, the introduction and the conclusion section of the reference papers, and the full text except the related work section of the new writing paper.

Table 2. The ROUGE scores of RWS-Cit-AIC, RWS-Cit-FULL, RWS-AIC-AIC, and RWS-AIC-FULL.

Metric	*RWS-Cit-AIC*	*RWS-Cit-FULL*	*RWS-AIC-AIC*	*RWS-AIC-FULL*
ROUGE-1 F-measure	0.45281	0.35413	0.40872	0.32830
ROUGE-1 Recall	0.55934	0.52379	0.44657	0.45015
ROUGE-1 Precision	0.39374	0.27321	0.38281	0.26313
ROUGE-2 F-measure	0.18008	0.12582	0.08849	0.06036
ROUGE-2 Recall	0.21846	0.18170	0.09892	0.08287
ROUGE-2 Precision	0.15788	0.09771	0.08148	0.04828

As shown in Table 2, *RWS-Cit-AIC* performs better than RWS-AIC-AIC. This is to say that using the citation document is better than using the main text of the references. Citations contain dense information than the main text of the references. *RWS-Cit-AIC* performs better than RWS-Cit-FULL, this is to say that using the abstract, the introduction, the conclusion section of the writing paper is better than using the full text of the wiring paper. Information in abstract and conclusion section is more concise than the main text.

VIII. CONCLUSIONS

This paper proposed a new approach to automatically generate the related work section for a writing paper by the following steps: 1) extract the citations of the reference papers from other papers to form a citation document, and the main text of the paper also forms a document for comparison; 2) construct feature graphs by extracting representative features; 3) generate Steiner tree from the feature graph; and, 4) extract the summaries that cover the tree as the summary. Experiments show that the proposed approach performs better than existing two baselines, and generating the related work section based on citations is better than generating the section based on the main texts of the reference papers.

Although the generated related work cannot completely represent the writer's own opinion and writing style, this work is significant because of the following three aspects: (1) it provides a new approach to generating comparative summarization; (2) the summary provides a draft of the related work for writer to revise, which is a way to save the writer's time for writing the related work from scratch that is usually time consuming in searching and read references, and more importantly, it can avoid missing relevant comments on related work; and, (3) this work verifies the general summarization method based on connotation and extension through citation [1].

The constraint of this approach is that the references should have citations. The more citation they have the better the effect of the proposed approach. For the new reference papers that have no citations, the direct summarization of the reference papers needs to be incorporated into the proposed approach.

ACKNOWLEDGMENT

Research was Sponsored by Natural Science Foundation of Jiangsu Province (BK20150862), NUPTSF(NY214166) and NUPTSF(NY215121).

REFERENCES

[1] H. Zhuge. Multi-Dimensional Summarization in Cyber-Physical Society. Morgan Kaufmann, 2016.

[2] S. Mohammad, B. Dorr, M. Egan, et al. Using Citations to Generate Surveys of Scientific Paradigms. Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, May 31 - June 5, 2009, Boulder, Colorado, USA. 2009:584-592.

[3] V. Qazvinian and D. R. Radev. Identifying Non-explicit Citing Sentences for Citation-based Summarization. ACL 2010, Proceedings of the, Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden. 2010:555-564.

[4] X. Wan, J. Yang, and J. Xiao. Manifold-ranking based topic-focused multi-document summarization. International Joint Conference on Artifical Intelligence. Morgan Kaufmann Publishers Inc. 2007:2903-2908.

[5] V. Qazvinian and D. R. Radev. Scientific Paper Summarization Using Citation Summary Networks. International Conference on Computational Linguistics. Association for Computational Linguistics, 2008:689-696.

[6] A. Elkiss, S. Shen, A. Fader, et al. Blind men and elephants: What do citation summaries tell us about a research article?. Journal of the American Society for Information Science & Technology, 2008, 59(1):51-62.

[7] C. D. V. Hoang and M. Y. Kan. Towards automated related work summarization. Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010: 427-435.

[8] Y. Hu and X. Wan. Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach. Conference on Empirical Methods in Natural Language Processing. 2014.

[9] D. H. Widyantoro and I. Amin. Citation sentence identification and classification for related work summarization. International Conference on Advanced Computer Science and Information Systems. IEEE, 2015:291 - 296.

[10] V. Qazvinian, D. R. Radev and A. Özgür. Citation Summarization Through Keyphrase Extraction. COLING 2010, International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2010, Beijing, China. 2010:895-903.

[11] V. Qazvinian, D. R. Radev, S. M. Mohammad, et al. Generating extractive summaries of scientific paradigms. Journal of Artificial Intelligence Research, 2014, 46(1):165-201.

[12] D. Wang, S. Zhu, T. Li, et al. Comparative document summarization via discriminative sentence selection. ACM Transactions on Knowledge Discovery from Data, 2009, 6(3):1963-1966.

[13] D. Wang, S. Zhu, T. Li, et al. Comparative Document Summarization via Discriminative Sentence Selection. ACM Transactions on Knowledge Discovery from Data, 2009, 6(3):1963-1966.

[14] R. Sipos and T. Joachims. Generating comparative summaries from reviews. ACM International Conference on Conference on Information & Knowledge Management. 2013:1853-1856.

[15] C. Shen and T. Li. Multi-document Summarization via the Minimum Dominating Set. COLING 2010, International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2010, Beijing, China. 2010:984-992.

[16] J. Li, L. Li and T. Li. Multi-document summarization via submodularity. Applied Intelligence, 2012, 37(3): 420-430.

[17] F. Jin, M. Huang and X. Zhu. A comparative study on ranking and selection strategies for multi-document summarization. COLING 2010, International Conference on Computational Linguistics, Posters Volume, 23-27 August 2010, Beijing, China. 2010:613-620.

[18] X. Huang, X. Wan and J. Xiao. Comparative News Summarization Using Linear Programming. The, Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, Usa - Short Papers. 2011:648-653.

[19] L. Zhang, L. Li, C. Shen, et al. PatentCom: A Comparative View of Patent Document Retrieval. Proceedings of the 2015 SIAM International Conference on Data Mining. 2015.

[20] H. Zhuge. Communities and emerging semantics in semantic link network: Discovery and learning. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(6): 785-799.

[21] H. Zhuge. The Knowledge Grid: Toward Cyber-Physical Society. World Scientific, 2012.

[22] H. Zhuge, Semantic linking through spaces for cyber-physical-socio intelligence: A methodology, Artificial Intelligence, 2011, 175: 988-1019.

[23] H. Zhuge, Interactive Semantics, Artificial Intelligence, 2010, 174: 190-204.

[24] H. Zhuge and B. Xu, Basic operations, completeness and dynamicity of cyber physical socio semantic link network CPSocio-SLN, Concurrency and Computation: Practice and Experience, 2011, 23(9): 924-939.

[25] H. Zhuge and J. Zhang, Automatically constructing semantic link network on documents, Concurrency and Computation: Practice and Experience, 2011, 23(9): 956-971.

[26] H.Zhuge and Y.Sun, The schema theory for semantic link network, Future Generation Computer Systems, 2010, 26(3): 408-420.

[27] C. D. Manning, M. Surdeanu, J. Bauer, et al. The Stanford CoreNLP Natural Language Processing Toolkit. Meeting of the Association for Computational Linguistics: System Demonstrations. 2014.

[28] M. Y. Day, T. H. Tsai, C. L. Sung, et al. A knowledge-based approach to citation extraction. IRI-2005 IEEE International Conference on Information Reuse and Integration, Conf, 2005. IEEE, 2005: 50-55.

[29] M. Y. Day, R. T. H. Tsai, C. L. Sung, et al. Reference metadata extraction using a hierarchical knowledge representation framework. Decision Support Systems, 2007, 43(1): 152-167.

[30] B. Powley and R. Dale. Evidence-based information extraction for high accuracy citation and author name identification. Large Scale Semantic Access to Content (Text, Image, Video, and Sound). LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 2007: 618-632.

[31] Y. Ding, G. Chowdhury and S. Foo. Template mining for the extraction of citation from digital documents. Proceedings of the Second Asian Digital Library Conference, Taiwan. 1999: 47-62.

[32] E. Cortez, A. S. da Silva, M. A. Gonçalves, et al. FLUX-CIM: flexible unsupervised extraction of citation metadata. Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries. ACM, 2007: 215-224.

[33] R. Mihalcea, C. Corley and C. Strapparava. Corpus-based and Knowledge-based Measures of Text Semantic Similarity. National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, Usa. 2006:775--780.

[34] J. Chen and H. Zhuge. Summarization of scientific documents by detecting common facts in citations. Future Generation Computer Systems, 2014, 32: 246-252.

[35] J. Chen and F. Wang. Expanding Citations in a Paper by Summarizing References Based on Co-Occurring Terms. The 10th International Conference on Semantics, Knowledge and Grids. China, 2014:108-111.

[36] Y. Yang and J. O. Pedersen. A Comparative Study on Feature Selection in Text Categorization. Fourteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 1997:412-420.

[37] R. Fraer. Minimum Spanning Tree. Program Development by Refinement. Springer London, 1999:151-165.

[38] C. Flick. ROUGE: A Package for Automatic Evaluation of sum-maries. The Workshop on Text Summarization Branches Out. 2004:25--26.

[39] D. Radev, T. Allison, S. Blairgoldensohn, et al. MEAD - a platform for multi-document multilingual text summarization. 2004.