IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727
PP 62-64
www.iosrjournals.org

PPS Gen: Learning-Based Presentation Slides Generation for Academic Papers

Parvez Shaikh
(*Computer Department, Rajarshi Shahu College of Engineering, University of Pune, India*)

Abstract: *Some rough structure for slide presentations from papers capable to save the author much time when organizing presentations. In this paper we investigate different perspective for academic papers slide generation. To write the slides from scratch takes a lot of time of presenter. They generally contain several sections like abstract, introduction, related work, proposed method, experiments and conclusions. To maintain uniqueness in preparing slides this idea is essential and unique. Each section from the academic paper is identified and is aligned to one or more slides. Every bullet point will be mapped with the slide heading point. Out of many sentences below that inside that heading sentences importance is calculated so as to keep those as it is in the slides.*

Keywords: *Abstracting methods, text mining*

I. Introduction

Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach

In this paper, author investigates a method for automatic related work generation. This method aims to generate a related work section for a multiple reference papers as input to target paper. Author proposed Automatic Related Work Generation system called ARWG to address this task. PLSA model is used to split the sentence set into different topic-biased parts. Regression model applies to learn the importance of the sentences. At last it employs an optimization framework to generate the related work section. Topic Model Learning is proposed using PLSA approach. Useful to handle mixture components as different words in a document may be generated from different topics.

Firstly, we use a PLSA model to group both sentence sets of the target paper and its reference papers into different topic-biased clusters. Secondly, the importance of each sentence in the target paper and the reference papers is learned by using two different Support Vector Regression (SVR) models. At last, a global optimization framework is proposed to generate the related work section by selecting sentences from both the target paper and the reference papers. Meanwhile, the framework selects sentences from different topic-biased clusters globally.

The importance scores for sentences in the target paper and the reference papers are assigned by using two SVR based sentence scoring models. Meanwhile, a topic model is applied to the whole set of sentences in both the target paper and reference papers. The sentences are grouped into several different topic-biased clusters. The sentences with importance scores and topic cluster information are taken as the input for the global optimization framework. The optimization framework extracts sentences to describe both the author's own work and background knowledge.

II. Literature Survey & Contribution

In proposed system abstract, introduction, conclusion and related work consider as input, we can make use of citation sentences for improvement.

Investigating Automatic Alignment Methods for Slide Generation from Academic Papers :

In this paper author proposed method for automatic generation of slide and also paper alignment. Four different alignment systems are used to compare which used in other alignment such as TF-IDF term weighting and query expansion. TF-IDF is nothing but simpler scoring mechanism. It is based only on the number of matched terms.

In previous methods they focus only on either aligning sentences to sentences or paragraphs to paragraphs. But our focus it to aligning slide regions which are usually bullets spanning at most a couple lines.
In some cases same information is assumed to be present in each document. This information may be presented in different way. But author is not able to make this assumption. Even they show that as much as half of the information in slide may not be present in the corresponding paper.

The general procedure for aligners is as below:

1. The token's TF-IDF score is calculated, the token's term frequency is the frequency of the token's stem in the region.
2. The SNoW tagger is used for and part-of-speech tagging and non content words are removed. Content words may be any token which is a noun, adjective, verb, adverb, or cardinal number.
3. Each token in the slide region is stemmed, in the case of aligners, query expansion is performed.
4. A score is calculated for each region in the target paper according to the scoring function

Result shows that query expansion does not improve performance of application and that TFIDF term weighting is inferior to a much simpler scoring mechanism based on the number of matched terms. In future performance can be improved by using terms in nearby regions to supplement both slide regions and paper regions.

Mining and Analyzing the Future Works in Scientific Articles:

In this paper, author use technique to mine the future works in scientific articles. This aims to

1. Provide an insight for future work analysis and
2. Facilitate researchers to search and browse future works in a research area.

First, used to study the problem of future work extraction and used regular expression based method for this. Secondly, four different categories are define for the future works by observing the data and investigate the multi-class future work classification problem. Third, we apply the extraction method and classification to a referred paper dataset in the computer science field and conduct a further analysis of the future works. Finally, we design a prototype system to search and demonstrate the future works mined from the scientific papers. No previous works have investigated the future work mining task. This problem is addressed very firstly in this paper.

For this author used two strategies as future work extraction and future work classification. Based on this results, author further analyze the future works. This allows finding more important research information. Four future work categories are define as "problem", "method", "evaluation" and "other". In the end, we design a system to search and rank the future works. The future work results can be displayed in different categories.

Hu and Wan (2014) introduced a system called ARWG to generate the related work section for the original target paper. They used supervised learning and an optimization framework to deal with the task. PDFlib2 is used to extract text and to detect their physical structures of paragraphs, subsections and sections ParsCit3 is used. It is possible to directly extract future work if that section is directly available. If not then we can extract the conclusion section.

Automatic slide presentation from semantically annotated documents:

This paper attempted to automatically generate slides from input documents annotated with the GDA tag set. GDA tagging can be used to encode semantic structure. The semantic relations extracted between sentences include grammatical relations, thematic relations and rhetorical relations. They first detect topics in the input documents and then extract important sentences relevant to the topics to generate slides.

A support system for making presentation slides (in Japanese):

This paper introduced a support system for making slides from technical papers. The inputs of the system are academic papers in LATEX format. The system calculates the weights of the terms in the paper using TF*IDF scores. Using the term weights, objects in the paper like sentences, tables etc. are also weighted and used to determine the number of objects for each section to generate the slides.

Automatic slide generation based on discourse structure analysis:

In this author proposed a method to automatically generate slides from raw texts. Clauses and sentences are considered as discourse units and coherence relations between the units such as list, contrast, topic-chaining and cause are identified. Some of clauses are detected as topic parts and others are regarded as non-topic parts. These different parts are used to generate the final slides based on the detected discourse structure and some heuristic rules.

III. Conclusion And Future Scope

After analyzing the above method for slide generation we suggest that a module capable of robustly filtering out unalienable slide regions. Also need to consider many factors, such as the contents, the venue information, the author information and the paper information.

In future it will be helpful if we apply the regression method to learn the importance scores of the sentences and use the integer linear programming (ILP) method to generate well-structured slides by selecting and aligning key phrases and sentences.

References

[1]. Yue Hu and Xiaojun Wan, *"PPSGen: Learning-Based Presentation Slides Generation for Academic Papers", IEEE Transactions On Knowledge And Data Engineering*, VOL. 27, NO. 4, APRIL 2015

[2]. M. Utiyama and K. Hasida, *"Automatic slide presentation from semantically annotated documents,"* in *Proc. ACL Workshop Conf.* Its Appl., 1999, pp. 25–30.

[3]. Y. Yasumura, M. Takeichi, and K. Nitta*, "A support system for making presentation slides,"* Trans. *Japanese Soc. Artif. Intell.*,vol. 18, pp. 212–220, 2003.

[4]. T. Shibata and S. Kurohashi, *"Automatic slide generation based on discourse structure analysis,"* in Proc. *Int. Joint Conf. Natural Lang. Process*., 2005, pp. 754–766.

[5]. T. Berg-Kirkpatrick, D. Gillick, and D. Klein, *"Jointly learning to extract and compress," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Lang. Technol.*, 2011, pp. 481–490

[6]. *Abstractive Summarization of Line Graphs from Popular Media* Charles F. Greenbacker Peng Wu Sandra Carberry Kathleen F. McCoy Stephanie Elzer, *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, pages 41–48, Portland, Oregon, June 23, 2011. c 2011 Association for Computational Linguistics

[7]. V. Qazvinian, D. R. Radev, S. M. Mohammad, B. J. Dorr, D. M.Zajic, M. Whidby, and T. Moon, *"Generating extractive summaries of scientific paradigms," J. Artif. Intell. Res.*, vol. 46, pp. 165–201, 2013.

[8]. Galanis, D. and Malakasiotis, P. (2008). AUEB at TAC 2008. *In Proceedings of the Text Analysis Conference, Gaithersburg, MD.*

[9]. Gillick, D. and Favre, B. (2009). *A scalable global model for summarization. In Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, pages 10–18, Boulder, CO.

[10]. Rani Nelken and Stuart M. Shieber. 2008. *Towards robust context-sensitive sentence alignment for monolingual corpora. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics.*

[11]. Amitay, E., and Paris, C., *"Automatically summarizing web sites: is there any way around it?"*, In Proceeding of the *9th International Conference on Information and Knowledge Management, McLean*, Virginia, November 2000, pp. 173-179.

[12]. André, E., Rist, T., Mulken, S. V., Klesen, M., and Baldes, S., "*The automated design of believable dialogue for animated presentation teams"*, In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, Embodied Conversational Agents, The MIT Press, 2000, pp. 220-255.