

Survey: A Learning Based Slide Generation from Academic Paper

Amritha S. Dev & Mrs.Amina Beevi A.

Department of Computer Engg.,Sree Buddha Engg. College For women,Elavumthitta,Pathanamthitta
amrithasahadev@gmail. comamilalu@gmail.com

Abstract: A presentation is the process of showcasing a subject to an audience. It is a form of lecture, demonstration, or speech that conveys information to listeners. Now in this digital era, presentation learning became an important way of expressing one's opinion. But making of these slides are tedious process that takes much time. A rough structure of paper in the form of presentation slide can save a lot of time of a presenter. This paper provides a novel system called PPSGen that help to automatically create slides from the academic paper. Here Support Vector Regression model (SVM) are used to find the important score of the sentences . The Integer Linear Programming (ILP) method is used to generate slide and detect the global and local phrases that can be used in slide. The slides produce the main topics of the paper and its respective main points. This paper shows a survey on PPSGen: Learning-Based Presentation Slides Generation for Academic Papers. The PPSGen has various advantages over baseline methods.

Keywords: Abstracting methods, text mining, Support Vector Regression model, Integer Linear Programming

I. Introduction

Presentation slides are the one of important way of learning. In this way of learning get the response from the listeners they may be students, employees or customer etc. Presenters create their slide by using the software tools like Microsoft Power-Point, Open Office etc. All these software presenters have to type the content into the slide then it will be time-consuming task. In this proposed method automatic slides are generated according to the sections in the paper i.e, titles in academic paper and corresponding relevant sentences from the same paper. It helps users in getting a rough structure of the academic paper.

In this survey, the proposed method PPSGen using SVM with some additional features for finding the importance of the sentences in the paper. The presentation slides are prepared with the ILP method with objective to pick the relevant sentences. This method has many advantages over the baseline methods.

II. Literature Survey

In M. Sravanthi et. al. [i] introduces the solution for reducing the effort of the presenter and help them in creating a structured summary of paper. It helps in creating slides for presentation with important point and all necessary figure etc. The vital points of the paper will mention. The Latex document is provided as the input and converted into xml format. The xml file will parse

and extract the information. A query specific extractors will help to summarize and generate the slides.

In Abu-Jbara et. Al. [ii] uses citation-based summarization approach to produce readable summarization of scientific papers. This approach removes the unimportant sentences, filters the sentences that are presentable and readable. In this paper, address solves the citation of sentences of the target paper belongs on. It is important to find aspect of the cited paper. In this method uses three steps such as, pre-processing, extraction, and post processing for citation-based summaries. The experiment produces better summaries than other several baseline summarization system.

In M. Utiyama et. Al. [iii] the authors explain the different method to automatically produce slides. The presentation system inputs a document annotated with the GDA tag-set, an XML tag-set which allows the machine to automatically infer the semantic structures underlying the raw document. The system takes important sections on the basis of conferences and semantic dependencies identified from the tags. This section selection also depends on interaction with the audience and help to create the presentation slides. Sentences important to the selected sections are extracted and paraphrased to form summary for the slide. The advantage of the system is that it is applicable to natural language and it is style-free or domain which can adopt different languages.

In T. Shibata et. al. [iv] finds a method to automatically produces slides from raw texts. Detection of Sentences and Clauses are identified between the units such as contrast, topic chaining, list and cause. In topic parts some of the clauses are detected and others are related to non-topic parts. These parts are used to produce the slides based on the detected discourse structure and some heuristic rules. The advantage of the paper are it will help in summarizing the document and semantically annotated document even if the paper are relatively long.

In B. Beamer et. al. [v] proposes method for automatic generation of slide and also paper alignment. Four different alignment systems, TF-IDF term weighting and query expansion are used to compare which used in other alignment. TF-IDF is similar to simpler scoring mechanism. It is based only on the number of matched terms and query aligner performance. This experiment shows 75% of accuracy.

The Mei et. al. [vi] proposes language modeling methods to combine features such as proximity and authority to predict the impact language model. The summarization test set is based on ACM SIGIR papers for greater impact. To identify the citation context and original content of a paper to produce impact-based summary authors suggest a language models. The paper helps to study about features such as authority and proximity into the estimation of language models. An impact-based summary is

used for facilitating the exploration of literature, it also helps to generate query.

In V. Qazvinian et. al. [vii], the proposed model summarizes single topic from the article and this summarized topic is further used to summarize the entire topic of the specified article. Clustering approach is using in this paper. The paper uses citation summaries and network analysis techniques which produce a summary of a single scientific article which helps in framework for future research on topic summarization. Corpus is built by extracting small clusters from the AAN data. Dependency Parsing (DP), Phrased Based Machine Translation (PBMT), Text Summarization (Summ), Question Answering (QA), and Textual Entailment (TE) are the clusters collected in this paper. Non overlapping contribution (fact) is used to show each item on the list. In this paper, for article summarization graph clustering method is used. Experimental result outperforms the current state of summarizing documents.

The Y. Yasumura et. al.[viii], introduced a support system for making slides from technical papers. The inputs of the system are academic papers in LATEX format. The system computes 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.

III.Comparison Table

| Paper | Method | Advantages | Future Works |
|---|--|---|--|
| SlidesGen: Automatic Generation of Presentation Slides for a Technical Paper Using Summarization[1] | A method of generating summarized version of paper with vital points. | Save time and effort. | Use of natural language processing technique and identify appropriate structure for slide to become appealing. |
| Automatic slide presentation from semantically annotated documents[3]. | A method of generating slides from semantically annotated document. | Audience interaction and can be applicable to natural language. | More synthetic evaluation and accommodate figures and tables. |
| Coherent citation-based summarization of scientific papers[2]. | Summarizing a scientific paper using its set of citation sentence called citation based summarization. | Produce better summary and increase readability. | Combination of auto-abstract variable length and emphases. |

| | | | |
|---|---|--|--|
| Automatic slide generation based on discourse Structure Analysis[4] | a method of automatically generating summary slides from a text | Generated slides are far easier to read. | Reduce non-topic parts in the slides, to obtain greater accuracy . |
|---|---|--|--|

IV. Conclusion

This survey paper proposes a novel system called PPSGen to generate presentation slides from academic papers. Sentence scoring model is based on SVR and uses the ILP method to align and extract key phrases and sentences for generating the slides. Experimental results show that our system will help in generating rough structure of the academic paper as slides with important points.

Presently the system help in generating slides of a single paper at a time. Further we can combine the different paper from same domain with same concepts can be merged and produce the slides, we can add different features in slides like hyperlinking the main concepts etc.

References

- i. M. Sravanthi, C. RavindranathChowdary and P. Sreenivasa Kumar, "SlidesGen: Automatic Generation of Presentation Slides for a Technical Paper Using Summarization", in *Proceedings of the Twenty-Second International FLAIRS Conference* (2009).
- ii. Abu-Jbara and D. Radev, "Coherent citation-based summarization of scientific papers," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Lang. Technol.-Volume 1, 2011*, pp. 500–509.
- iii. M. Utiyama and K. Hasida, "Automatic slide presentation from semantically annotated documents," in *Proc. ACL Workshop Conf. Its Appl.*, 1999, pp. 25–30.
- iv. 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.
- v. B. Beamer and R. Girju, "Investigating automatic alignment methods for slide generation from academic papers," in *Proc. 13th Conf. Comput. Natural Lang. Learn.*, Jun. 2009, pp. 111–119.
- vi. Mei and C.Zhai, "Generating impact-based summaries for scientific literature," in *Proc. ACL*, vol. 8, pp. 816–824, 2008.
- vii. V. Qazvinian and D. R. Radev, "Scientific paper summarization using citation summary networks," in *Proc. 22nd Int. Conf. Comput. Linguistics-Volume 1*, Aug. 2008, pp. 689–696
- viii. 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
- ix. S. M. A. Masum, M. Ishizuka, and M. T. Islam, "Auto-presentation: A multi-agent system for building automatic multi-modal presentation of a topic from World Wide Web information," in *Proc. IEEE/WIC/ACM Int. Conf. Intell. Agent Technol.*, 2005, pp. 246–249.
- x. H. P. Luhn, "The automatic creation of literature abstracts," *IBM J. Res. Develop.*, vol. 2, pp. 159–165, 1958.
- xi. S. M. A. Masum and M. Ishizuka, "Making topic specific report and multimodal presentation automatically by mining the web resources," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, 2006, pp. 240–246