International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Review on Presentation Slides Generation for Scientific Papers

Vishakha N. Pawar[1], Manisha M. Naoghare[2]

[1]Student, Master of Engineering, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Sinner

[2]Assistant Professor, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi, Sinner

Abstract: ***This paper discusses a way for slides generation without human intervention from a text, studying the generation of presentation slides from a scientific paper and also examines the demanding task of incessantly creating presentation slides from scientific papers. The produced slides can be used as a rough copy to help moderators setup their orderly slides in a swift manner. This paper introduces a novel system called PPSGen to assist everyone create such slides. The system uses the revert procedure to conclude the Score of the sentences in an paper and then uses the whole number Integer Linear Programming (ILP) system to create well-ordered slides by selecting and changing key vocabulary and sentences. Results are based on a test game plan of 200 arrangements of papers and slides assembled on the web, display that our proposed PPSGen structure can make slides with improved quality. A client study also exhibits that PPSGen has palpable benefit over baseline methods.***

Keywords: Abstracting methods, text mining, Support Vector, Regression (SVR), ILP, Classification

1. Introduction

Presentation slides are a helpful means to communicate and share information and convey key-messages across the listeners at educational as well as professional meetings. The research-presenter uses slides to share information in a tidy and articulate format. The research-presenter has plentiful programming tools to lend a hand in setting up the slides, including Microsoft Power-Point and Open Office. Various tools available help research-presenter in setting up the backgrounds and outline of the presentation; apply various themes; however they do not help research-presenter in selecting the substance for the slides. The conventional tools therefore require a lot of deal, in terms of efforts and time, from the research-presenter. In this work, a policy is proposed for making presentation slides for scientific papers. The main goal is to generate draft slides for the research-presenters so as to minimize their efforts and time in manage up presentation slides. Spontaneous papers have a mostly reliable structure and enclose a couple of sections like introduction, system overview, related work, proposed work, experiments, evaluation and conclusions. Different people create presentations in their own different ways; however the person generally adjusts slides in the order as mentioned in the various areas of the paper.

PPSGen maps every area of paper to one or more slides with an emblematic slide having a title and a few point wise explanations. These explanations may be included in some visual sequences. Our policy attempts to produce run-of-the-mill draft slides to assist persons in setting-up their final slides. Programming slides for scientific papers proves to be a remarkably complex task. Present systems focus on substance like sentences from the paper to construct the slides. Slides can be secluded into different parts. Each part means a particular point and contains topics, which are essential to one another.

In this study, we put forward the PPSGen system to make pre-composed presentation slides for scientific papers. In our structure, the significance of each sentence in a paper is figured out by using the Support Vector Regression (SVR) model with different cooperative components. The presentation slides for the paper are then twisted by using the Integer Linear Programming (ILP) model with complicatedly ordered board limits and objectives to accept and alter key vocabulary and sentences. It has been evaluated on a test course of 200 paper-slides sets, which display that our methodology can produce slides with superior quality over the standard frameworks.

2. Framework

Unlike reviews, we provide an extra general idea on the overall procedure of automatic slide generation, which is summarized in figure 1. In this paper, it review current developments and analyze future open directions in slide generator which works automatic. The main contribution of this survey are as follows 1) Sentence scoring and slide generation is mentioned in a evidently ordered, hierarchical manner and the interlink between these components is shown 2) To scrutinize the state of the art, every job involved in slide generation is alienated into sub-process and various types of approaches to the sub-process are mentioned. The intrinsic worth and drawback of the various approaches are also mentioned into next section.

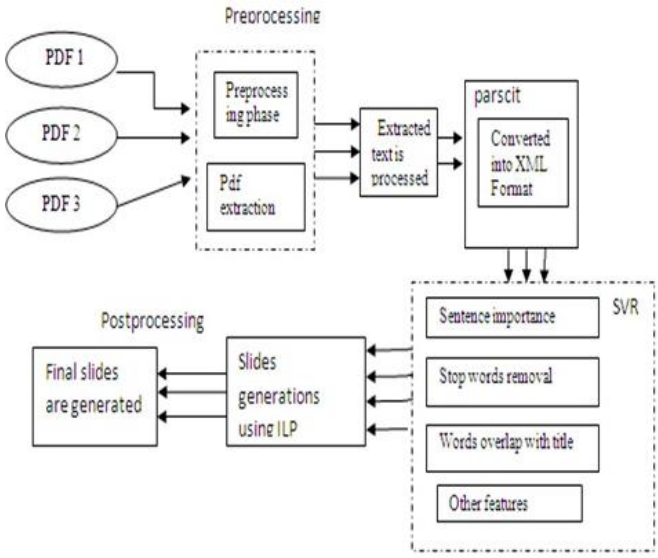

Figure 1: Generic framework for automatic slides generator for scientific papers

A. Sentence selection (scoring)

Presentation slides are completed with bullet points and corresponding explanation sentences. These sentences are selected using some particular process, so that they can be used to display on the presentation slides. Various methods work in a different way to select the sentences from scientific papers. This results in assortment of most suitable and significant sentences. One more concept i.e. Summarization is used to select particular summary and produce slides using this explanation available in summary, but the slides are prepared with the help of available concept which contains only sentences but not key phrases aligned to the sentences.

Y. Yasumura et al [2] proposed a hold for creating slides from technical papers. The academic papers were used as inputs to the system in Latex format. The weight of the terms is calculated in the paper using TF*IDF scores.

All objects in the paper like sentences, tables etc. are also weighted using weight of the terms. Based on the weights of the objects, sentences can be extracted. This extraction can be used for each section in the paper. After extraction, generate the slides using a slide composition template, which can be edited by the users.

Luhn et al [10] has outlined a new technique to create abstract automatically using modern electronic data processing devices i.e. Summarization. Experts of scientific papers and magazines that offer the idea of conventional abstracts have been created entirely by involuntary means. Machine uses the Statistical information derived from word frequency and distribution to work out a relative measure of significance, first for individual words and then for sentences. Sentences scoring highest in significance are extracted and printed out to become the "auto-abstract".

Another method is projected in [11] for formation of an automatic index from technical documents. The author told that if the sentence falls with comparatively high occasion at various fixed position within the paragraph, it is a simple stuff to have the machine select the sentence and trace it for compiling an abstract or for extracting the vocabulary to form an index. Examination of 200 paragraphs confirmed the fact that in 85% of the paragraphs the topic sentence was the first one and in 7% of the time it was the last sentence. A trouble-free way to select the topic sentence would be a program the machine to select the first and the last sentence of the paragraph.

B. Slide Generation

After first step of scoring the sentences, the most significant sentences are segregated using some particular method. Well-structured slides are created by Integer Linear Programming method is by selecting and aligning key phrases and sentences. Bullet points are created by Key phrases and sentences relevant to the phrases are placed below the bullet point. In order to extort the key phrases, Open NLP library is applied to implement chunking to the sentences and noun phrases are extracted as the candidate key phrases. Two kinds of phrases are defined: global phrases and local phrases. Global Phrase is unique phrase found in an article, and a local phrase means a global phrase in a particular section. For example, "SOFTWARE" is a global phrase of the paper, while its appearances in various sections are measured different local phrases. "SOFTWARE {Introduction}" and "SOFTWARE {Our Proposed Method}" denote different local phrases, and they symbolize the appearances of "SOFTWARE" in Sections 1 and 4 of the paper, respectively. Few local phrases correspond to a global phrase that appears in various sections. Since a crucial phrase is always used in various sections, a global phrase that corresponds to more local phrases should be regarded important and more likely to be selected. Thus, the bullet points are created by local phrases directly for different sections and use the global phrases to tackle the main differences between all unique phrases. Stemming and Stop word removal process are used for the phrases. Moreover, the noun phrases which appear only once in the paper are discarded.

Automatic slide presentation is created using GDA tag set in [1] for. The Slides are created using the semantically annotated documents. Attempted to automatically generate slides from input documents annotated with the GDA tagset.1 GDA tagging can be used to encode semantic structure. The semantic relations include grammatical relations such as subject, thematic relations such as agent, patient, and rhetorical relations such as cause and elaboration. They first detect topics in the input documents and then extract important sentences relevant to the topics to generate slides.

For making presentation slides a TF*IDF method is presented in [2]. A technical paper presentation slides is made using support system. This system provides functions that allocate slides to each and every section and puts objects on a slide. A paper as a latex document is main input to this system; the next input to the system is the number of slides user wants and keywords of the paper. Initially the system converts a paper from a text document into an XML document. The XML document can comprise information of a paper such as ID number and term weights. Next, the weights of term is calculated in the system in the document by the TF*IDF method. The document consists of various

objects such as figures, tables and sentences and they are weighted by term weights. Using the weight of the objects and slide merging template, the system decides how many slides are assigned to each section.

A new method is developed by Shibata and Kurohashi in [3], 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. The final slides are generated based on the detected discourse structure and some heuristic rules.

3. Literature Survey

The various methodologies are discussed in paper [1] by the authors to automatically generate slides.
Inputs for the system are as follows:
1. A document annotated with the GDA tag-set
2. XML tag-set.

It allows the machine to automatically conclude the semantic structures essential the raw document. The automatic system takes an important topic on the basis of semantic dependencies and conferences recognized from the tags. The selection of the topic depends on communication with the audience and dynamic adaption of the presentation is the result of it. Relevant Sentences to the chosen topic are extracted. Itemized summary is formed for the slide by paraphrasing. Heuristics are applied for creating layout and paraphrasing. GDA tag-set is independent of the domain and style of document. GDA tag-set is applicable for a numerous natural languages. Due to above characteristics of GDA Tag-set, the reported system is also independent of domain and the adaption of different languages is of ease.

A support system which produces presentation from latex document is introduced in paper [2].

The system provides following functions:
1) Assign slides to each section
2) Put objects on a slide.

Input to this system is as follows:
1) Technical paper as a latex document.
2) Number of slides user want to make.
3) Keywords of the paper.

Initially the system converts from a tex document into a XML document. The information such as term weights, ID number of paper, etc. is included in the XML documents. In next step the system uses TF*IDF method to calculate weights of term in the document. The objects such as figures, tables, sentences, etc. in the document are weighted on term weights. The slides are assigned to each section depending on the weights of the objects and slide symphony template. The composition is done by system depending upon above template. In First Step the texts are retrieved depending on user's query which is most similar to it. These retrieved texts are used to generate the slides automatically is shown in paper [3]. Then in next step the system converts the printed text into verbal languages. In later step the system feed the converted text to a speech synthesis engine as printed text are not appropriate, because aberrant speech might be formed due to difficult words or long compound nouns, which are inappropriate for speech synthesis. The paraphrasing technique is used to convert written texts into spoken texts. Later this output is used as input into speech synthesis, but the negative aspect of this slide generation is that it contains non-topic parts along with topic parts.

In paper [4] author the draw near of obtaining a set of rules for producing presentation sheets by using machine learning techniques to numerous pairs of technical papers and their presentation sheets composed from world wide web. As a first step in this paper, a method is introduced for aligning technical papers and presentation sheets and the method is based on Jing's method which uses a Hidden Markov Model.

A Digital Library which consists of only published documents by the researchers is shown in paper [5]. The research factories of the researchers are converted into written document and then slide presentation. As these research credentials are very exclusive and include very valuable information, so as a substitute of to refer these two documents separately, it is superior to align and present such presentation document pairs together. So such alignments between these two are done in digital library. The three major system components of the Slide-Seer DL:
1) Resource discovery
2) Fine-grained alignment
3) User interface.

In [6] proposes method for automatic generation of slide with paper alignment. TF-IDF term weighting and query expansion are used to evaluate which used in other alignment. TF-IDF is analogous to simpler scoring mechanism. It is based only on the number of matched terms and query aligner performance. This experiment shows 75% of accuracy.

In Paper [7] author explained a new agent based scheme where in the user gives queries as input. The information regarding query is searched on internet in the background. Images are added to the output slides by the system. The system worked on different techniques like web page parsing, web data fetching and summary extraction. Each operation was done by agents. Web data fetching and parsing is done by some specific algorithm. In the post processing phase, MPML scripts were created and the output slides were produced in HTML and Javascript formats and the topics were elaborated in different headings by agent characters.

In paper [8], the data is available and accessible in large amount which is Internet-based resources and robust nature of Web pages. Due to this nature, the task of information retrieval is getting difficult and complex. Some key areas focused in this paper are as follows:
1) Agent based autonomous system.
2) Automatic report to presentation (ARP), with the idea of autonomous information service emerging as the result of integration among natural language processing,

3) Web intelligence.
4) Character-based agent interaction.

The system, ARP follows the process of fetching a set of Web-pages; and then parsing, summarizing affect-senses and correlates information extracted from those and finally automatically building a report on a topic and search phrase given by a user.

The Important aspect of the paper is identified using citation-based summarization, text written by numerous researchers in paper [9]. Previously, Extraction was the main feature to work on this problem. Meanwhile, the facility of the producing summaries has not been given that much significance. For example, readability, diversity, cohesion and ordering is included in summary. This leads to noisy and confusing summaries. In this work, they present a way to produce readable and cohesive citation-based summaries. The experiments show that the proposed approach gives outstanding performance in several baselines in terms of both extraction quality and fluency.

4. Conclusions

We propose a novel system called PPSGen to create presentation slides from scientific papers. Sentence scoring model is skilled based on SVR and the ILP method is used to align and extract key phrases and sentences for creation of slides. The proposed strategy is mainly useful to create enormously better slides than habitual routines are shown by experimental results. Presently, our system generates slides based on only one given paper, but in future additional information such as other relevant papers and the citation information can be used to improve the generated slides. Also Graphical images can be added to the slides from scientific papers.

5. Acknowledgements

It is a great pleasure to acknowledge those who extended their support, and contributed time and psychic energy for the completion of this seminar work. The authors would like to thank the publishers and researchers for making their resources available. We also thank the college who served as sounding board for both contents and programming work. For their valuable and skilful guidance, assessment and suggestions from time to time improved the quality of work in all respects. We would like to take this opportunity to express my deep sense of gratitude towards them, for their valuable contribution in completion of this paper. We would like to thank our colleagues and friends who helped me directly a indirectly to complete this paper. Finally our special thanks to our family members for their support and cooperation during this Work.

References

[1] M. Utiyama and K. Hasida, "*Automatic slide presentation from semantically annotated documents,*" in Proc. ACL Workshop Conf. Its Appl., 1999, pp. 25–30.
[2] 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.
[3] 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.
[4] T. Hayama, H. Nanba, and S. Kunifuji, "*Alignment between a technical paper and presentation sheets using hidden Markov model*," in Proc. Int. Conf. Active Media Technol., 2005, pp. 102–106.
[5] M.Y. Kan, "SlideSeer": *A digital library of aligned document and presentation pairs*," in Proc. 7th ACM/IEEE-CS Joint Conf. Digit. Libraries, Jun. 2006, pp. 81–90
[6] 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.
[7] 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/ACMInt. Conf. Intell. Agent Technol., 2005, pp. 246–249.
[8] 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.
[9] A. 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.
[10] H. P. Luhn, "*The automatic creation of literature abstracts*," IBM J. Res. Develop., vol. 2, pp. 159–165, 1958.
[11] P. B. Baxendale, "*Machine-made index for technical literature: an experiment*," IBM J. Res. Develop., vol. 2, no. 4, pp. 354–361,1958.
[12] V. Qazvinian and D. R. Radev, "*Identifying non-explicit citing sentences for citation-based summarization*," in Proc. 48th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2010, pp. 555–564. 1096 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015.

Author Profile

Vishakha N Pawar received the B.E. degree in Computer Engineering from N.D.M.V.P's College of Engineering in 2005.She worked with various industries in Software Development. She is currently pursuing Masters Degree in Computer.

Manisha M. Naoghare is working as Assistant Professor in Sir Visvesvaraya Institute of Technology.