

LEARNING-BASED PRESENTATION SLIDES GENERATION AND IEEE FORMAT CREATION FOR ACADEMIC PAPERS

Miss. Priti Patil¹ Mr. R. V. Patil², Mr. H. T. Ingale³

¹ME Scholar, E&TC Engg. department, GF's Godavari College of Engineering, Jalgaon-India

²Prof of E&TC dept, E&TC Engg. department, GF's Godavari College of Engineering, Jalgaon-India

³Head, E&TC dept, E&TC Engg. department, GF's Godavari College of Engineering, Jalgaon-India

ABSTRACT

We uses the backslide procedure to determine the Noteworthiness Score of the sentences in a paper and then uses the Integer Linear Programming Algorithms system to create well-organized slides by selecting and adjusting key expressions and sentences. Evaluated result based on a certain set of 200 arrangements of papers and slide assemble on the web displays in our proposed structure of PPSGen can create slides with better quality and quick. Paper talks about a technique for consequently getting outline slides from a content, contemplating the programmed era of presentation slides from a specialized paper and also examines the challenging task of continuous creating presentation slides from academic paper. The created slide can be used as a draft to help moderator setup their systematic slides in a quick manners. This paper introduces novel systems called PPSGen to help moderators create such slide. A customer study also exhibits that PPSGen has obvious advantage over baseline method and speed is fast for creations.

Keyword : - Artificial Support Vector Regression (SVR), Integer Linear Programming (ILP), Abstract methods, texts mining, Classification etc....

1. INTRODUCTION

Presentations slide are an effective medium to convey and share information's and to deliver key-message across the audience at professional and educational meeting. The research-presenter makes use of slide to share information's in an order and lucid formats. The research-presenter has number of programming tool to assists him in setting up the slide, including Microsoft Power-Point, Open Office. Such tools help researchers in setting up the themes and outline of the presentation; however they does not help researcher in selecting the contents for the slide. The traditional tools thus require a lot of investments, in terms of time and effort, for the researchers. In this work, a strategy is proposed for making presentations slide for academic paper. Aim is to generate draft slides for the research-presenters so as to reduce their time and effort in setting the presentation slide. Intuitive paper has a relatively consistent structure and contains a couple of different section like introductions, system overviews, related work, propose strategies, examinations and conclusions.

In our structures, the importance of each sentence in a paper is figure out by using the Support VectorRegression models with the help of different accommodating component. The presentations slides for the papers are then created by using the Integer Linear Programming models with the help of complicatedly arranged targets limits and objective to pick and alter key expression and sentence. It has been examined on test courses of 200 paper-slide set, which demonstrate that our methodology can create slide with better quality over the standard framework on Internet. Using the ROUGE tools and the Pyramid Appraisals, the slides made by our methods can enhance ROUGE score and Pyramid score. Additionally, in light of customer focused, our slide can get higher rating score by human judge in both substances and structure point.

2. EARLIER WORK

Unlike the review, we give more general overviews on the overall processes of automatic slides generation, which is outlined in figure. It reviews a recently developments and analyze future open direction in automatic slide generators. The key contribution of this survey is as follow,

- 1) The sentence scoring and slides generations is discussed in a clear organized, hierarchical manner and the interlink between these component is shown
- 2) To examine the state of the art, each task involved in slide generation is dividing into sub-process and various categories of approaches to the sub-process are discusses. The merits and limitation of the different approach are summarized.

A. Sentence selection (scoring)

Presentations slide are made up of bullet point and corresponding sentences so that they are refined. The selection of these sentences is done via using some specific methods, so that they can be display on the presentations slide. Different methods worked differently for selecting the sentence from academic paper. This result in selection of most appropriate and relevant sentence. One more concept i.e. Summarization is to select specific summary and generate using this summary, but with the help of this concept slides prepare which contains only sentence but not key phrases aligned to the sentence.

Luhn et al [1] has outlined a technique for automatic creations of abstract using modern electronic data processing devices i.e. Summarizations.

Y. Yasumura et al [5] introduced a supports for making slide from technical paper. The inputs of the system are academic papers in Latex formats. The system calculates the weight of the term in the papers using scores.

In [11], another method is proposed for creation of an automatic index from technical documents. The author says that if the sentence falls with relatively high occurrence at some fixed position within the paragraph, it is an easy matter to have the machines select the sentences and record it for compiling an abstracts or for extracting the vocabulary to form an index.

B. Slides Generations

After the tasks of score the sentence, the most important sentence are selected using some specific methods. Integer Linear Programming methods is used to generate well-structured slide by selecting and aligning key phrases and sentences. In order to extract the key phrase, chunking implement by the Open NLP libraries is applied to the sentences and noun phrase are extracted as the candidates key phrases. Two kinds of phrases are defined: global phrase and local phrase. Any unique phrase in articles is a global phrase, and a local phrase means a global phrase in a particular sections. So a global phrase that appears in different sections can correspond to a few local phrases. Since an important phrase is always used in more different sections, a global phrase that corresponds to more local phrases should be regarded to be more important and more likely to be selected. Thus, the local phrases are used to generate the bullet points directly for different sections and use the global phrases to address the importance difference between different unique phrases. All the phrases are stem and stop word are removed. Moreover, the noun phrases which appears only once in the paper are discarded.

GDA tag sets is used in [1] for the automatic slide presentation from semantically annotated documents.

In the paper [3] exploits a method to automatically generate the 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.

3. LITERATURE SURVEY

In paper the author discussed the various methodologies to automatically generate the slides. The given reported presentations system is getting input a document annotate with the GDA's tag-sets, a XML tag-sets which has allow the machines to automatically infers the semantics structure underlying the raw documents. The system picks up important topics on the basis of semantics dependency and conferences identified from the tag. These topic

selections also depend on interactions with the audiences and further lead to dynamic adoptions of the presentations. Sentences relevant to the selected topics are then extracted and paraphrased to form an itemized summary for the slides. Same heuristics are applied for paraphrase stage and layouts. Since the GDA tag-set is independent of the domains and style of document and applicable for a variety of natural languages, the reported system is also domain independent and easy to adapt to different languages.

In paper [3] the slides are generated automatically, but before generation of slides the texts are retrieved which is most similar to the users query.

In paper[4] author the approach of obtaining a set of rules for generating presentation sheets by applying machine learning technique to many pair of technical paper and their presentations sheets collected from worldwide webs.

In paper [5] A Digital Library consists of only published documents by the researchers. The research works of the researchers are transmitted into written document and slide presentations. The three major system components of the Slide-Seer DL: 1) the resource discovery, 2) the fine-grained alignment and 3) the user interface.

In the paper [6] Authors introduce a support systems for making presentations slides from a latex documents. This system provides functions that assign slides to each section and put objects on a slide. The XML document can include information of a paper such as ID number and term weights. Next the systems calculate weights of term in the document by the method.

In paper [8] Due to the availability and accessibility of large Internet-based resources and robust natures of Web pages, the task of information retrieval is becoming more challenging and complex. Agent based autonomous system, automatic reports to presentation (ARP), with the notion of autonomous information service emerging as the result of integration among natural language processing, Web intelligence, and character-based agent interactions are the key areas focused in this paper.

In paper [9] citation-based summarization, text written by several researchers is made use of to identify the important aspects of a target paper. The fluency of the produced summaries has not been given that much importance. If the summary which includes diversity, readability, cohesions, and ordering of the sentences have not been thoroughly considered. This led to noisy and confusing summaries. [18] In this work, they present an approach for producing readable and cohesive citation-based summaries.

4. PROPOSED SYSTEM

In this case, we propose a system to automatically generate slides that have good structure and content quality from academic papers. The architecture of our system will be shown in Fig. 4. We use the SVR-based sentence scoring model to assign an importance score for each sentence in the given paper, where the SVR model is trained on a corpus collected on the web. Then, we generate slides from the given paper by using ILP.

1.1 Sentence Importance Assessment:

In our proposed PPSGen system, sentence importance assessment is one of the two key steps, which aims to assign an importance score to each sentence in the given paper. The score of each sentence will be used in the slides generation process. In this study, we introduce a few useful features and propose to use the support vector regression model to achieve this goal.

1.2 Support Vector Regression Model:

Here, we briefly introduce the SVR model [23]. We need to predict the importance score of each sentence for sentence selection in slides generation. The reason why we use the SVR model instead of the classification model is that the regression score is finer to be used for sentence selection than the coarse binary category.

<http://arnetminer.org>

<http://www.pdflib.com/>

<http://aye.comp.nus.edu.sg/parsCit/>

<http://www.foolabs.com/xpdf/>

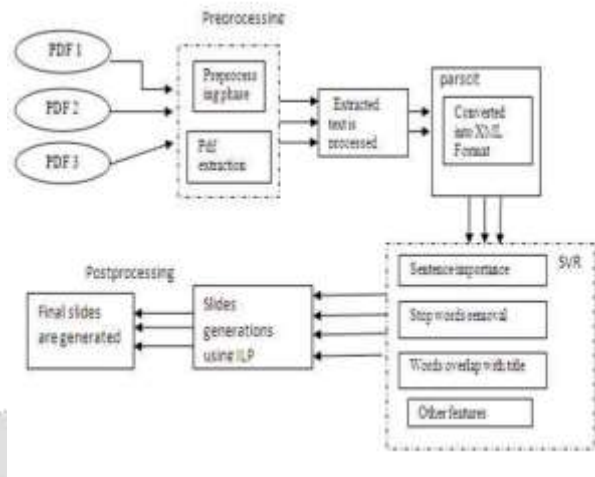


Fig -1: Diagram of PPSGen

5. SYSTEM SOFTWARE DESCRIPTION

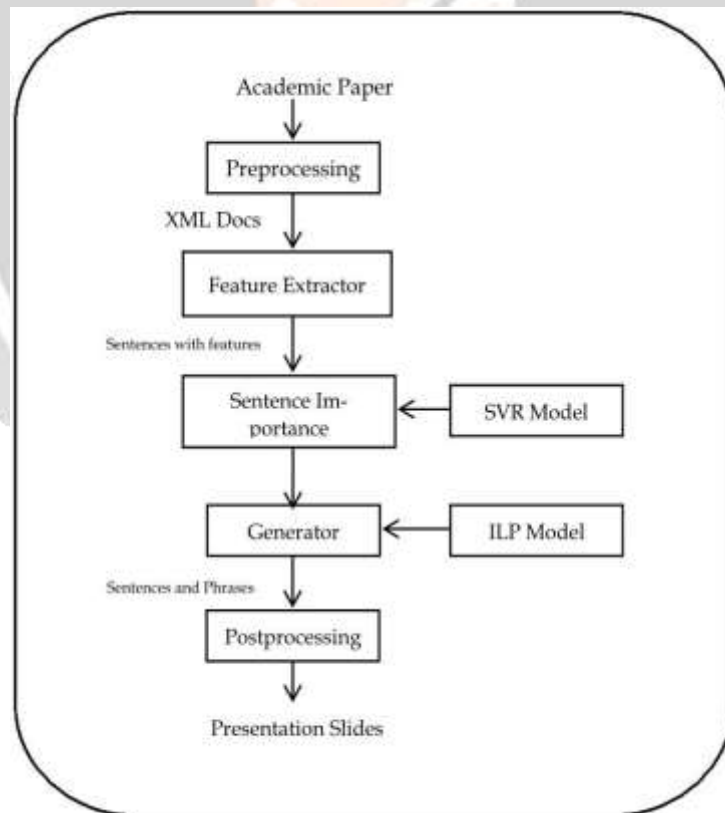


Fig -2: Name Flowchart for Slide Generation

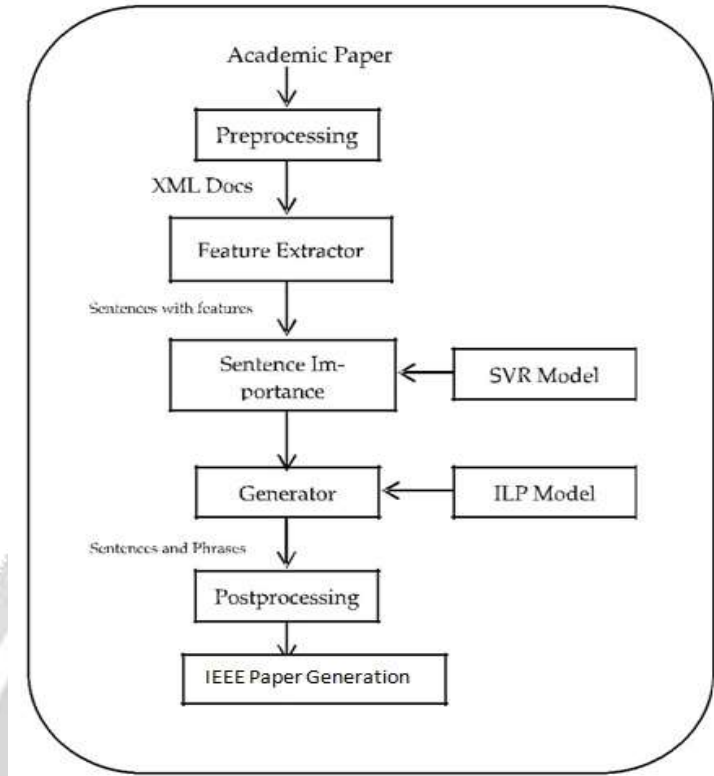


Fig -3: Flowchart for IEEE format Creation



Fig -4 : Front End UI for User

Figure 4 shows the Front End User Design Which Illustrate user for the services the application will be providing the user home is the common way point for all the services for the user.

6. CONCLUSION

This paper proposes a novel system called PPSGen to generate presentation slides from academic papers. We train a sentence scoring model based on SVR and use the ILP method to align and extract key phrases and sentences for generating the slides. Experimental results show that our method can generate much better slides than traditional methods. In future work, we will improve our system by using both text and graphical elements in the paper and make slides more comprehensible and vivid. When dealing with the graphical elements, we need to identify the graphical elements in the paper first. The relationship between the text elements and the graphical elements also needs to be identified. We need to know which sentences are most relevant to a graphical element and which graphical elements should be selected to generate the slides. We can use rule-based methods or machine learning based methods to solve the above problems. Then we can simply attach the tables and figures we select to the most relevant sentences in the slides. In this paper, we only consider one typical style of slides that beginners usually use. In the future, we will consider more complicated styles of slides such as styles that slides are not aligned sequentially with the paper and styles that slides have more hierarchies. We will also try to extract the slide skeletons from the human-written slides and apply these slide skeletons to the automatic generated slides. Furthermore, our system generates slides based on only one given paper. Additional information such as other relevant papers and the citation information can be used to improve the generated slides. We will consider this issue in the future.

6. REFERENCES

- [1]. H. P. Luhn, "The automatic creation of literature abstracts," IBM J. Res. Develop., vol. 2, pp. 159–165, 1958.
- [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]. N. Ahmed, Joint Detection Strategies for Orthogonal Frequency Division Multiplexing, *Dissertation for Master of Science, Rice University, Houston, Texas*. pp. 1-51, Apr. 2000.
- [4]. 3GPP RP-040461, Study Item: Evolved UTRA and UTRAN, December 2000
- [5]. 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.
- [6]. Parsaee, G.; Yarali, A., "OFDMA for the 4th generation cellular networks" in Proc. IEEE Electrical and Computer Engineering, Vol.4, pp. 2325 - 2330, May 2004.
- [7]. J. D. Poston and W. D. Horne, "Discontinuous OFDM considerations for dynamic spectrum access in idel TV channels," in Proc. IEEE DySPAN, 2005.
- [8]. T. Hayama, H. Nanba, and S. Kunifuji, "Alignment between a technical paper and presentations sheets using hidden Markov model," in Proc. Int. Conf. Active Media Technol., 2005, pp. 102– 106.
- [9]. S. M. A. Masum, M. Ishizuka, and M. T. Islam, "Auto-presentations: 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. Intel. Agent Technol., 2005, pp. 246– 249.
- [10]. 3GPP R1-050971, "R1-050971 Single Carrier Uplink Options for EUTRA: IFDMA/DFT-SOFDM

Discussion and Initial Performance Results "http://www.3GPP.org,Aug 2005

- [11]. IEEE P802.16e/D12,'Draft IEEE Standard for Local and metropolitan area networks-- Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems', October 2005
- [12]. S. R. Baig, F. U. Rehman, and M. J. Mughal, "Performance Comparison of DFT, Discrete Wavelet Packet and Wavelet Transforms in an OFDM Transceiver for Multipath Fading Channel," 9th IEEE International Multitopic Conference, pp. 1-6, Dec. 2005.
- [13]. M.Y. Kan, "SlideSeer": A digital library of aligned documents and presentation pairs," in Proc. 7th ACM/IEEE-CS Joint Conf. Digit. Libraries, Jun. 2006, pp. 81– 90
- [14]. R. Rajbanshi, Q. Chen, A. Wyglinski, G. Minden, and J. Evans, "Quantitative comparison of agile modulation technique for cognitive radio transceivers," in Proc. IEEE CCNC, Jan. 2007, pp. 1144–1148.
- [15]. V. Chakravarthy, Z. Wu, A. Shaw, M. Temple, R. Kannan, and F. Garber, "A general overlay/underlay analytic expression for cognitive radio waveforms," in Proc. Int. Waveform Diversity Design Conf., 2007.
- [16]. R. Dilmirghani, M. Ghavami, "Wavelet Vs Fourier Based UWB Systems", 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp.1-5, Sep. 2007.
- [17]. M. Weeks, *Digital Signal Processing Using Matlab and Wavelets*, Infinity Science Press LLC, 2007.
- [18]. V. Chakravarthy, Z. Wu, M. Temple, F. Garber, and X. Li, "Cognitive radio centric overlay-underlay waveform," in Proc. 3rd IEEE Symp. New Frontiers Dynamic Spectrum Access Netw., 2008, pp. 1–10.
- [19]. R. Mirghani, and M. Ghavami, "Comparison between Wavelet-based and Fourier-based Multicarrier UWB Systems", IET Communications, Vol. 2, Issue 2, pp. 353-358, 2008.
- [20]. [20]V. Chakravarthy, X. Li, Z. Wu, M. Temple, and F. Garber, "Novel overlay/underlay cognitive radio waveforms using SD-SMSE framework to enhance spectrum efficiency—Part I," IEEE Trans. Commun., vol. 57, no. 12, pp. 3794–3804, Dec. 2009.
- [21]. X. Li, R. Zhou, V. Chakravarthy, and Z. Wu, "Intercarrier interference immune single carrier OFDM via magnitude shift keying modulation," in Proc. IEEE Global Telecomm. Conf. GLOBECOM , Dec. 2009, pp. 1–6.
- [22]. 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.
- [23]. Xue Li, , Vasu D. Chakravarthy, , Bin Wang, and Zhiqiang Wu, "Spreading Code Design of Adaptive Non-Contiguous SOFDM for Dynamic Spectrum Access" in IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 5, NO. 1, FEBRUARY 2011
- [24]. 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.

- [25]. 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.
- [26]. 1096 *ieee transactions on knowledge and data engineering*, VOL. 27, NO. 4, APRIL 2015.
- [27]. Published in: *Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, *International Conference on* Date Added to IEEE Xplore: 16 March 2017 INSPEC Accession Number: 16743975 DOI: [10.1109/ICACDOT.2016.7877699](https://doi.org/10.1109/ICACDOT.2016.7877699) Publisher: IEEE

