SLIDEGen: Approach to Automatic Slide Generation

Vishakha N. Pawar[1], Prof. K. N. Shedge[2]

Student, Master of Engineering, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology,

Chincholi, Sinner, Nashik, Maharashtra, India. [1]

Head of Department, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Chincholi,

Sinner, Nashik, Maharashtra, India. [2]

ABSTRACT: Presentations are one of the most familiar and valuable way of communicating the general idea of the work to spectators. Given a technical paper, automatic generation of presentation slides minimizes the pains of the presenters and it also helps in creating an ordered summary of the paper. In this, we look into a very tough task of automatically generating slides for presentation of academic papers or research papers. The generated slides can be presented to prepare their proper slides in an easier way. It first uses a model called SVR model (Support Vector regression model). System use Support Vector regression model for sentence significance assessment by checking importance score of each sentences in an academic paper. Another model named ILP model is used. After getting importance score for each sentence from academic word document with the help of SVR and ILP method is used to generate fine draft slides with content as key sentences. A data mining approach for automating the generation of presentation slides from an academic article is presented. Initially, the system is trained using a large dataset to learn the intricacies involved in how humans do the task of slide generation. The input to the proposed scheme is a technical article.

KEYWORDS: Data mining, Powerpoint Generation, Automatic Slide generation, Research Papers.

I. INTRODUCTION

PRESENTATION slides have been a mainstream what's more, powerful intends to display and exchange data, particularly in meetings and gatherings, a moderator takes the aid of slides to exhibit his work in a systematic way (pictorial).In recent years with the accessibility of numerous software tools like Microsoft Power- Indicate and OpenOffice presenter etc., offer specialists some help with making ready their slides. In any case, these apparatuses just guide them inside the arranging of the slideshow, however certainly not inside substance. Regardless you will require arbitrators long to create slides compose this slides. In this particular do the job, most of us propose to her some sort of technique for consequently creating presentation slides for scholarly papers. We all mean to naturally produce very much organized photo slides and also give such draft slides as a premise to lessen the moderators chance and exertion while setting up their last speech slideshow par Scholastic papers dependably have a comparative design. They for the most part contain a few areas like conceptual, presentation, related work, proposed system, analyses and conclusions. Despite the fact that presentation slides can be composed in different routes by distinctive moderators, a moderator, particularly a learner, dependably adjusts slides consecutively to the paper areas while setting up the slides. Every segment is adjusted to one or even more slides and another glide ordinarily carries a concept along with a couple sentences. These kinds of sentences might be incorporated into some visual cues. Our strategy endeavours to create write slideshow in the average sort said previously mentioned the ones groups to set up their last slides.

II. RELATED WORK

In M. Utiyama. Al. [1] the authors explains 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.

Yasumura et al. [2] presented an emotionally supportive network for making slides from specialized papers. The inputs of the framework are scholarly papers in LATEX position. The framework computes the weights of the terms in the paper utilizing TF-IDF scores.

Shibata et al., [3] described a method for the generation of presentation slides from an article by the analysis of the discourse structure of the article. A clause and sentence was considered as a discourse unit by the system and the important coherence relations such as contrast, list, additive, elaboration etc were extracted and analyzed.

In paper, T. Hayama et al., [4] the approach of obtaining a set of rules for generating presentation sheets by applying machine learning techniques to many pairs of technical papers and their presentation sheets collected from world wide web.

M.Y. Kan [5] displayed Digital Library framework which consists of just distributed records by the specialists. The research works of the researchers are transmitted into composed document and, slide presentation.

In B. Beamer et. al. [6] 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.

Shaikh Mostha Al Masum et al., [7] elaborated a new agent based scheme where in the user could give queries as input. In the background the system collected information about the query by searching the internet. Images could also be added to the output slides by the system.

Mistsuru Ishizuka et al., [8] Due to availability and accessibility of large Internet-based resources and robust nature of Web pages, the task of information retrieval are becoming more challenging and complex. Agent based autonomous system, automatic report 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 interaction are the key areas focused in this paper.

Sravanthi et al., [9] elaborated a system to auto generate presentation slides from a research manuscript. Here also, the starting point of slide generation is from the raw LaTeX source of the research manuscript. After the inference of the logical structure from the article, each section was categorized to fall under Introduction, Related Works, Model, Experiments and Conclusion respectively.

The process of automatic slide generation in this technique involved the use of QueSTS summarizer [10]. Graphical elements could also be extracted from the article by the system and the slides were built.

III. PROPOSED ALGORITHM
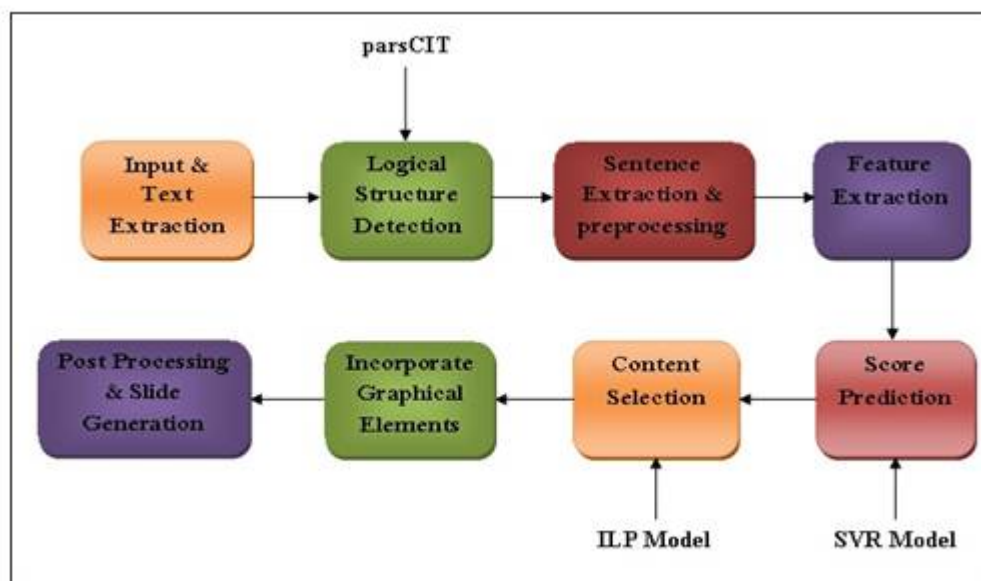
A. Module Description

1. Input and Text Extraction:
In this step, the input is specified as a Portable Document Format file. The system extracts the text contents of the research manuscript and is saved as a text file.

2. Logical Structure Detection:
The system infers the logical hierarchy of sections, subsections, subsubsections by using parsCit[13] . The input to parsCit is the text contents of the manuscript. The output of parsCit is an XML file with information on the logical structure of the article.

3. Sentence Extraction and Preprocessing:
With the help of the inferred logical structure, the system extracts the individual sentences with information on their positional details i.e., to which section/ subsection/ subsubsection they belong. After this process, common preprocessing techniques like stemming and stop word removal are done on all sentences.

4. Feature Extraction:
The system considers each individual sentence and extracts its features. This forms the test dataset and is analyzed in the next phase.

5. Score Prediction:
The test dataset is analyzed by the SVR Model and an importance score is predicted for each sentence. The score indicates the level of importance of a sentence. The score value lies between 0 and 1. Higher the score, higher the probability for that sentence to be selected to the slides.

6. Content Selection:
For selecting the contents to be included to the slides, the system uses an Integer Linear Programming Model with a robust objective function having well defined constraints. In this module, the system collect global phrase in a particular section is known as a local phrase. Instead of just placing some extracted sentences as a summary for a section/ subsection, the system identifies the most important key phrases using the ILP Model since it would have more emphasis on the actual concept portrayed by the particular section.

7. Incorporate Graphical Elements:
Graphical elements can also be incorporated to the output slides and the final presentation is produced in either TeX or PPT formats.

8. Post Processing :
In the post processing phase, MPML scripts were generated and the output slides were formed in HTML and Javascript formats and the topics were explained to different headings by agent characters.

B. *Proposed Algorithm:*

Input: PDF file
Output: Slide Generated with graphical element and text.
Process:
1: Read PDF File.
2. Apply Parscit to detect their physical structures of paragraphs, sections and sections
3: Token generation and apply stemmer.
4: Remove Stop words and apply post tagging.
5: Calculate sentence scoring using NLP.
6: Add Image into Slide (See detail in Algorithm 2)
7. Add scoring sentences and image using ILP method
8: Generate the slide.

IV. RESULTS

We choose user evaluation technique to find out the result of our method. we did evalution by asking the authors of the papers to rate the presentation generated by our system. We collected source pdf of 10 papers and generated presentations for them. They were given to each one of the authors of the corresponding papers. The authors were asked to answer the following questions by giving a rating on a scale of 1 (bad) to 10 (good) for the presentations.

Q1 : The level of information covered in the created presentation?
Q2 : The level of coherence in the slides?
Q3 : The level of presentation for a good starting point to prepare final presentation?
Q4 : All images from paper is included ?
Q5 : Overall satisfaction level with the presentation?
Tables are handled in a similar fashion, but with a few notable differences

USER	Q1	Q2	Q3	Q4	Q5
U1	10	10	09	10	09
U2	09	09	09	09	09
U3	08	07	08	07	08
U4	09	09	10	10	10
U5	10	10	10	10	10
U6	09	09	10	10	09
U7	08	07	09	08	09
U8	09	08	08	08	09
U9	09	09	08	08	09
U10	10	09	10	08	08
AVG	9.1	8.7	9.1	8.8	9.0

Evaluation Results of Presentation

V. CONCLUSION AND FUTURE WORK

The proposed system is 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. Graphical images can be added to the slides from scientific papers.

Future Enhancement

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. Although a lot of work has been done, but still there is a scope for the more efficient .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.

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.13thConf. Comp. 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. M. Sravanthi, C. R. Chowdary, and P. S. Kumar, "SlidesGen: Automatic generation of presentation slides for a technical paper using summarization," in Proc. 22nd Int. FLAIRS Conf., 2009, pp. 284–289.
10. H. P. Luhn, "The automatic creation of literature abstracts," IBM J. Res. Develop., vol. 2, pp. 159–165, 1958.

BIOGRAPHY

Mrs. Vishakha Nilesh Pawar, received the B.E. degree in Computer Engineering from N.D.M.V.P's College of Engineering in 2005.Appeared for M.E Computer from SVIT, Chincholi at Pune University. Currently Professor in Guru Gobind Singh Polytechnic, Nashik. She has over 12 years of experience in teaching and Software industry. Her Research Interest are Data Mining, ERP, Graphics, etc

Prof. K. N. Shedge is working as Head of Department, Computer Engineering in Sir Visvesvaraya Institute of Technology, Chincholi.