ASAR Automatic PPT Generation Using Machine Learning

Shivam Kumar Singh[1], Rahul Sharma[1], Prof S.V Shinde[2]

[1]BE, Department of Computer Engineering, AISSMS College of Engineering, Pune,
Maharashtra, India

[2]Assistant Professor, Department of Computer Engineering, AISSMS College of Engineering, Pune,
Maharashtra, India

ABSTRACT

In most of the areas for sharing information, slide presentation plays an important role. The slides for the presentation are traditionally prepared using various tools. The traditional way of presenting slides is labor-intensive. Labor-intensive nature leaves scope for human-errors. In this paper, Research observation is enforcing the automated PPT creation from multi-documents of different extensions based on input query or title that formulate extraction of valuable information source and model a presentation view to automating slide creation using integer linear programming (ILP) method to generate well-structured slides by selecting and aligning key phrases and sentences. This will eventually help in reducing a great amount of the presenter's time and efforts. The proposed system works on natural language processing (NLP) rules to classify data for the desired slides. Preparing slides manually consume more time. The drawbacks of the traditional way lead to the need for an intelligent system. The intelligent system needs to be capable of generating slides with minimum human interference. The existing automatic tools fail to fetch the graphical elements from a given input. Hence the paper proposed an Automatic Slide Generation System. The proposed system fetches the graphical elements as well as text from a document. The proposed system is more reliable than the existing system.

Keywords: Classification, NLP, Support Vector Regression (SVR), ILP, Slide Generation, Modules

I. INTRODUCTION

As one of the most important ways of transferring information and knowledge, presentations provide users to discuss and exchange ideas together. Presentations now play an important role to promote understanding of presenters ideas in many fields such as education, research and business. Many university lecturers use Web service to store, browse, and share presentation slides that are used in their lectures. Although powerful tools for Inserted: slide composition have been developed and Web service has been widely used to share slides, they have a problem with preparing many lecture slides to help their students understand the contents. Lecturers need to prepare slides that can promote understanding. But how to make the content of the slide fine-grained and highly structured is still a problem to address[4]. The traditional tools such as Microsoft PowerPoint and OpenOffice Impress fail to provide any functions of handling presenters' intentions. For example, the logical design of the slide structure and the interrelations between the objects. In presentation slides composition, it is necessary for presenters to consider topics, items, and their relations. Audiences cannot understand the presentation until they see the consistent discussion points in the slide structure. key idea is to construct a Language function from the objective function which is called as the primal objective function and the corresponding constraints, by introducing a dual set of variables. Automatic slides generation for

multiple documents is a very challenging task[1]. Current methods generally extract objects like sentences from the file to construct the slides. In contrast to the short summary extracted by a summarization system, the slides are required to be much more structured and much longer. Slides can be divided into an ordered sequence of parts. Each part addresses a specific topic and these topics are also relevant to each other. Generally speaking, automatic slide generation is much more difficult than summarization. Slides usually not only have text elements but also graph elements such as figures and tables.

Our work can be summarized as follows:
- ✓ We propose a framework to support the presentation slide design activities.
- ✓ We integrate the content-based element search method to help user reuse the existing materials,
- ✓ We provide different methods to support the presentation design in different stages.
- ✓ We preliminarily conduct experiments to evaluate the effectiveness and the usefulness of the proposed framework.

In this paper, we investigate the problem of support for presentation slide design. Unlike previous methods, we consider the role users play in the presentation structure design and content generation. The users may select topics from a statistic list to design the structure of the presentation, then input or search the contents for each topic unit, with the help of the element search method, the user can reuse the elements returned slides. Next, users can allow component into one slide, and select or specify the layout of the slide according to their preference. Based on the proposed framework, a prototype system with a user-friendly interface is designed. The experimental evaluation of real presentation slide design and generation demonstrates the effectiveness and usefulness of the proposed framework.

II.   EXISTING SYSTEM

Manual systems put pressure on people to be correct in all details of their work at all times, the problem is people are not perfect, many of us wishes were. With the manual systems, the level of service is dependent on individuals and puts a requirement on management to run training continuously for staff to keep them motivated and to ensure they are following the correct procedures[3]. It can be all too easy to accidentally switch details and end up with inconsistency in data entry. It has the effect of not only causing problems with customer service but also making information unable be used for reporting or finding trends in data discovery. Reporting and checking data is robust can be timely and expensive. It is often an area where significant money can be saved by automation.

A. Traditional Approach: Writing manual PPT

Many tools help the presenter to generate the slides. These tools only help them in the formatting of the slides, but not in the content. It still takes presenters much time to produce the slides from the available material. The traditional tools thus require a lot of investment, both in terms of time and efforts. Duplication of data entry may happen. The task of generating presentation slides from one or many written materials is both tedious and time-consuming[2]. It takes more effort and physical space to keep track of paper documents, to find information and to keep details secure. When mistakes are made or changes or corrections are needed, mostly a manual editing must be completely done again rather than just updated. With manual or partially automated systems information often has to be written down and copied or entered more than once. Systemization can reduce the amount of duplication of data entry.

B. Disadvantages of Existing system
- Focuses on summarization of the content.
- It does not generate slides and it does not consider graph elements.

- Sentence importance assessment is not up to quality standards.
- Most of the documents are in PDF format so many documents cannot be processed.

III. PROPOSED METHODOLOGY

We propose a system to automatically generate slides that have good structure and content quality from multiple documents. The architecture of our system is shown in Figure 1. 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. Automatic slides generation for academic papers is a very challenging task. Current methods generally extract objects like sentences from the file to construct the slides. In contrast to the short summary extracted by a summarization system, the slides are required to be much more structured and much longer. Slides can be divided into an ordered sequence of parts. Each part addresses a specific topic and these topics are also relevant to each other. Generally speaking, automatic slide generation is much more difficult than summarization.

Slides usually not only have text elements but also graph elements such as figures and tables. But our work focuses on the text elements only.
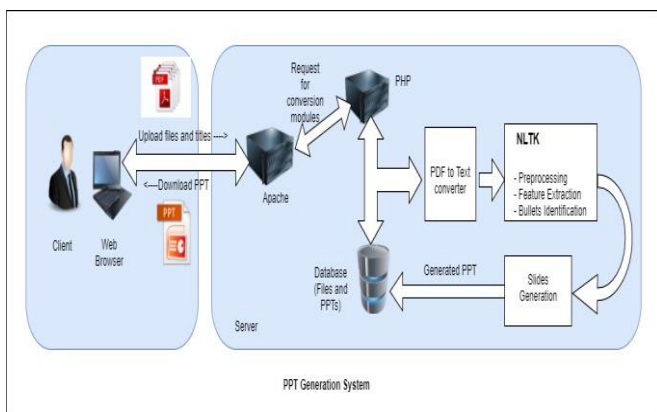

Figure 1. Proposed System (ASAR)

In this study, we propose a novel system to generate well-structured presentation slides for academic papers. In our system, the importance of each

sentence in a paper is learned by using the Support Vector Regression (SVR) model, and then the presentation slides for the paper are generated by using the Integer Linear Programming (ILP) model to select and align key phrases and sentences. Documents always have a similar structure. They generally contain several sections. Although presentation slides can be written in various ways by different presenters, a presenter, especially for a beginner, always aligns slides sequentially with the paper sections when preparing the slides. Each section is aligned to one or more slides and one slide usually has a title and several sentences. These sentences may be included in some bullet points[7]. Our method attempts to generate draft slides of the typical type mentioned above and helps people to prepare their final slides.

A. The overall system design consists of following modules:

(a) Accept multiple input files.
(b) Pre Processing using NLTK.
(c) Feature extraction using NLP.
(d) Sentence classification using SVR.
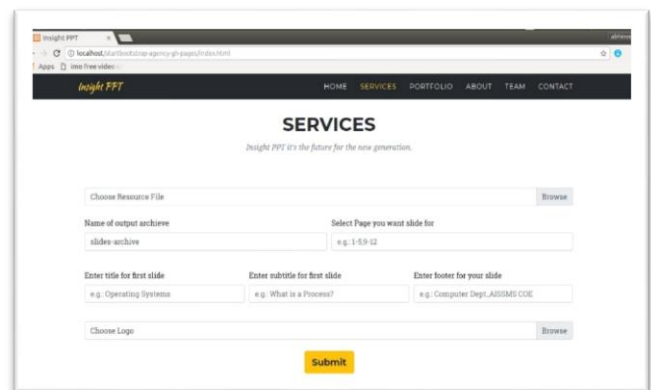(e) Generate slide using ILP.

IV. SYSTEM DESIGN

Figure 2. GUI of ASAR

A. Graphical User Interface:

The proposed system provides a better, efficient way to generate powerpoint presentation based powerful GUI for helping people to feed data using this form.

After submitting the result be downloaded in the form of zip format.

B. Algorithm for preprocessing:

Step 0: Start

Step 1: Read string

Step 2: divide the string into records on space and store in a vector V

Step 3: Remove Special Symbols

Step 4: Identify Stopwords

Step 5: Remove Stopwords

Step 6: Identify Stemming Substring

Step 7: Replace Substring to desire String

Step 8: Concatenate Strings

Step 9: stop

C. Mathematical model:

Mathematical model of the proposed system is stated below. The system S is represented as:

$S = \{I, P, N, F, X\}$

Input Browse Dataset

$I = \{I1, I2, I3, ..., in\}$

Where, I is a set of number of documents.

Preprocessing Method

$P = \{p1, p2, p3, ..., pn\}$

Apply token generation, stemming, removing stop words and post tagging.

p1,p2,p3…, pn are the number of preprocessing process.

NLP Model

$N = \{n1, n2, n3, ..., nn\}$

Where, N is represented as a set of NLP model In NLP method, calculate the score of the sentence.

n1, n2, n3,…..,nn are the number of NLP model process.

Finding Image

$F = \{f1, f2, f3, ..., fn\}$

Where, F is represented as a set of Finding Image

f1, f2, f3, …., fn are number of Finding Image process.

ILP Method

$X = \{x1, x2, x3, ..., xn\}$

Where X is represented as a set of ILP processing

In ILP method, important sentences are added to the slide along with the images.

x1, x2, x3,….., xn are number of ILP processing process.

V. RESULT AND DISCUSSION

This system automatically generates the slides from the research paper. This section shows the result obtained the proposed system. The Figure 1 shows the Average Pyramid Scores of TF-IDF, GDA and another technique for the generated presentation slides. Overall research on the paper shows that proposed method can generate slides with better quality than the baseline methods. Therefore, our slides are considered a better basis for preparing the final slides.
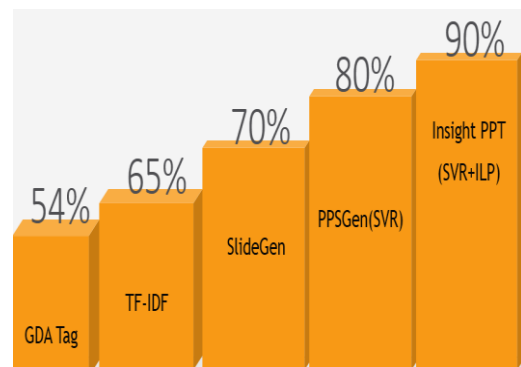
Figure 3. Performance evaluation

VI. CONCLUSION

The system is proposed for automatically generating the slides from the pdf. The system generates the slides which include text and graphical element. The graphical elements along with the text data make the generated slides look more comprehensible and vivid. The system initially finds the graphical elements of each page from the paper and after that system stores the image to map data. NLP method is used for sentence scoring and the ILP method is used for slide generation which contains key phrases and the relevant sentences. But in future multiple PPT with graphics feature can be provided

VII. REFERENCES

[1]. R. Jha, A. Abu-Jbara, and D. Radev, "A system for summarizing scientific topics starting from keywords", ACM Compute. Surv, vol. 40, no. 3, p. 8, 2013.

[2]. 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. 500509.

[3]. 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. 284289.

[4]. 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. 8190.

[5]. M. Utiyama and K. Hasida, "Automatic slide presentation from semantically annotated documents", in Proc. ACL Workshop Conf.Its Appl., 1999, pp. 2530.

[6]. 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.

[7]. Ektaa Meshram, D. A. Phalke, "Technique for Generating Automatic Slides on the basis of Paper Structure Analysis", in IJIRSET, Vol. 5, Issue 6, June 2016

[8]. MTech Scholar S. V Shinde, Prof. S.Z. Gawali, "Identification of keywords and phrases in text Document and sensing a word for document retrieval and ranking: First Review", in IJAIEM Volume 3, Issue 5, May 2014, ISSN 2319 4847.