IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

Research in SLIDEGens: Approach to Automatic Slides Generation

Miss. Autade Dhanshri P[1], Prof. Raut S.Y[1]

Pravara Rural Engineering College, Loni, India[1]

Abstract: In this paper, presentation are a standout amongst the most widely recognized and viable methods for conveying the review of a work to the gathering of people. We execute consequently producing presentation slides for scholastic papers. The created presentation slides can be used as drafts to offer the mediators some help with setting up their formal slides quickerly. A novel structure called PPSGen is proposed to address this endeavor. It first utilizes the relapse technique to take in the sentence significance evaluation in an insightful paper by utilizing SVM, and after that enterprises the Integer linear programming (ILP) strategy to make all around sorted out slides by selecting and modifying key expressions and sentences. Evaluation deciding results on a test set of 200 arrangements of papers and slides accumulated on the web demonstrate that our proposed PPSGen system can make slides with better quality. A customer study is in like manner outlined to exhibit that PPSGen has two or three clear inclinations over standard schedules.

Keywords: Support Vector Regression (SVR), ILP, bullet points.

I. INTRODUCTION

Presentation slides have been a prominent and powerful intends to present and exchange data, particularly in scholarly gatherings. The analysts dependably make utilization of slides to display their work pictorially on the meetings. There are numerous virtual products, for example, Microsoft Power- Point and Open Office to help specialists set up their slides. Be that as it may, these devices just help them in the designing of the slides, yet not in the substance. Regardless it takes moderators much time to compose the slides starting with no outside help. In this work, we propose a strategy for consequently creating presentation slides for scholastic papers. We expect to naturally create very much organized slides and give such draft slides as a premise to diminish the moderators' opportunity and exertion while setting up their last presentation slides.

Scholarly papers dependably have a comparative structure. They by and large contain a few areas like conceptual, presentation, related work, proposed technique, investigations and conclusions. Despite the fact that presentation slides can be composed in different courses by various moderators, a moderator, particularly a apprentice, dependably adjusts slides consecutively to the paper segments while setting up the slides. Every segment is adjusted to one or more slides and one slide typically has a title and a few sentences. These sentences might be incorporated into a few visual cues. Our strategy endeavors to create draft slides of the commonplace sort said above and people groups to set up their last slides.

Programmed slides era for scholarly papers is a extremely difficult undertaking. Current techniques by and large concentrate objects like sentences from the paper to develop the slides. Rather than the short rundown separated by a outline framework, the slides are required to be much more organized and any longer. Slides can be separated into a requested grouping of parts. Every part addresses a particular subject and these points are likewise important to each other. As a rule, programmed slide era is considerably more troublesome than outline. Slides more often than not have content components as well as chart components such as figures and tables. Be that as it may, our work concentrates on the content components as it were.

In this study, we propose a novel framework called PPSGen to produce very much organized presentation slides for scholarly papers. In our framework, the significance of every sentence in a paper is found out by utilizing the support vector regression (SVR) model with various helpful components, and after that the presentation slides for the paper are created by utilizing the whole integer linear programming (ILP) model with extravagantly planned target capacity and imperatives to choose and adjust key expressions and sentences.

Tests a test set of 200 paper-slides sets show our trategy can create slides with preferable high quality over the standard routines. Utilizing the ROUGE toolbox furthermore, the pyramid assessment, the slides created by our strategy can improve ROUGE standing in addition to pyramid standing. In addition, in light of a client review, our own 35mm slides can get better standing results by individual family court judges inside both. substance in addition.

II. RELATED WORK

A. SlidesGeneration

The Programmed slides era for scholastic papers stays far under-examined these days. Few concentrates straightforwardly look into on the theme of programmed slides era. Utiyama and Hasida [1] endeavored to naturally produce slides from info records explained with the GDA tagset.1 GDA labeling can be utilized to encode semantic

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

structure. The semantic relations incorporate syntactic relations, for example, subject, topical relations, for example, operator, persistent, and explanatory relations for example, cause and elaboration. They first identify subjects in the info reports and after that concentrate vital sentences pertinent to the points to produce slides.

Yasumura et al. [2] presented an emotionally supportive network for making slides from specialized papers. The inputs of the framework are scholastic papers in LATEX design. The framework figures the weights of the terms in the paper utilizing TF-IDF scores. Utilizing the term weights, objects in the paper like sentences, tables and so on are additionally weighted. Taking into account the weights of the items, the framework chooses the quantity of the articles like sentences to be extricated for every segment in the paper and after that produce the slides utilizing a slide sythesis layout which can be altered by the clients. Shibata and Kurohashi [3] proposed a technique to consequently produce slides from crude writings. Provisions and sentences are considered as talk units and soundness relations between the units, for example, list, contrast, topic chaining what's more, cause are recognized. Some of statements are distinguished as subject parts and others are viewed as non-point parts. These diverse parts are utilized to create the last slides taking into account the recognized talk structure and a few heuristic tenets.

Hayama et al. [4], Kan [5] and Beamer and Girju [6] concentrated on the issue of adjusting specialized papers and presentation slides. Hayama et al. utilized a variety of the Hidden Markov Model (HMM) to adjust the content in the slides to the undoubtedly segment in the paper, which likewise utilized the extra data of titles and position holes. Kan [5] connected an adjusted most extreme similitude technique to do the monotonic arrangements what's more, prepared a classifier to identify slides which ought not be adjusted. Beamer and Girju [6] looked at and assessed four distinctive arrangement techniques that were consolidated by strategies for example, TF-IDF termweighting and question extension.

Masum et al. [7], [8] proposed a framework named programmed report to presentation (ARP) which develops a point particular report and a presentation on a point or search queries given by a client. The framework recovers site pages significant to the disambiguated inquiry utilizing different web indexes. Headings and content pieces are separated from site pages and used to manufacture the report. A presentation is created by arbitrarily selecting up to five lines from every head-content tuple, two lines from the main, one in the center and the other two lines from the end of the content piece.

B. Scientific Article Summarization

Agarwal et al. [8] presented an unsupervised methodology to the issue of multi-record exploratory article rundown. The information is a rundown of papers refered to together inside the same source article. The key purpose of this methodology is a subject based grouping of parts separated from each cocited article. Yeloglu et al. [9] thought about four distinctive methodologies for multi-report investigative articles rundown: MEAD, MEAD with corpus particular vocabulary, LexRank and W3SS.

C. Document Summarization

ntered Gillick et al. [11], [12] and Berg-Kirkpatrick et al. [13] presented also, embraced an ILP technique in light of the idea of "ideas" which are really bigrams. Every idea (bigram) has a weight w. Every sentence is considered to comprise of a set of ideas and the ILP approach intends to boost the weights of the ideas secured by a choice of sentences. Woodsend and Lapata [14] additionally embraced techniques based on ILP to concentrate outline. The item capacity of the ILP model consolidates the significance of the bigrams in the synopsis' sentences, the striking nature of the parse tree hubs of the synopsis' sentences and a unigram dialect model which punishes sentences containing words that are likely to show up in synopses.

III. PROPOSED SYSTEM

Problem Definition:

Generation is very different from traditional rundown and experimental synopsis. These people simply select numerous sentences through the documents, whilst glides creation is really a lot more complex. Our recommended approach not only decides on many important essay sentences but also the actual words corresponding towards essay sentences. Soon after selecting paragraphs as well as key phrases, we can easily develop well-structured slideshow

A. Overview

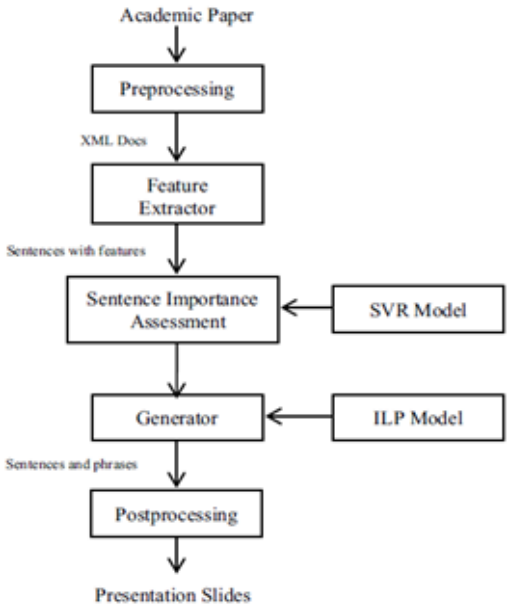


Fig. 1: A System Architecture

B. Support Vector Regression

Support Vector Machine can likewise be used as a regression system, sustaining every one of the primary characteristics which characterize the particular algorithm (maximal edge). The Support Vector Regression (SVR) employs identical guidelines as the SVM regarding class, together with only some modest distinctions. Above of all, because output is a genuine number it turns out to be exceptionally hard to anticipate the current data, which has

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

unbounded conceivable outcomes. With regards to regression, any border regarding ceiling (epsilon) is defined in think on the SVM which would have previously inquired from the dilemma. However aside from this particular reality, there's moreover. an even more complex motive, your criteria is usually more complex thus to get taken in factor.. Be that as it may, the main idea is always the same: to minimize error, individualizing the hyperplane which expands the edge, remembering that part of the mistake is endured. Support Vector Machine can be connected to grouping issues as well as to the instance of relapse. Still it contains all the primary elements that portray most extreme edge calculation: a non-linear capacity is inclined by direct learning machine mapping into high dimensional piece incited highlight space. The limit of the framework is controlled by parameters that don't rely on upon the dimensionality of highlight space.

C. Stop Word removing

Stop words will be words which are sifted through before or in the wake of planning of ordinary tongue data (content). Any gathering of words can be picked as the stop words for a given reason. A large portion of the most oftentimes utilized words as a part of English are futile in content mining these words are called stop words.

D. Stemming

After removing high frequency words, an indexing methodology tries to conflate word variants into the same stem or root utilizing a stemming algorithm. Case in point, the words "thinking", "thinkers" or "thinks" might may be reduced to the stem "think". In information retrieval, gathering words having the same root under the same stem (or indexing term) may increase the achievement rate when matching documents to a query.

Stop word removing

1 Define stopwords in separate file.
2 Traverse through the each sentence.
3 Check whether stopword present in sentence or not.
4 If stopword found, then that stopword get removed.
5 Else iterate through next line.
6 Continue until all lines are scan.

Step for stemming

1: Gets rid of plurals and -ed or -ing suffixes"
2: Turns terminal y to i when there is another vowel in the stem
3: Maps double suffixes to single ones:-ization, -ational, etc.
4: Deals with suffixes, -full, -ness etc.
5: Takes off -ant, -ence, etc.
6: Removes a final –e

IV. RESULT ANALYSIS

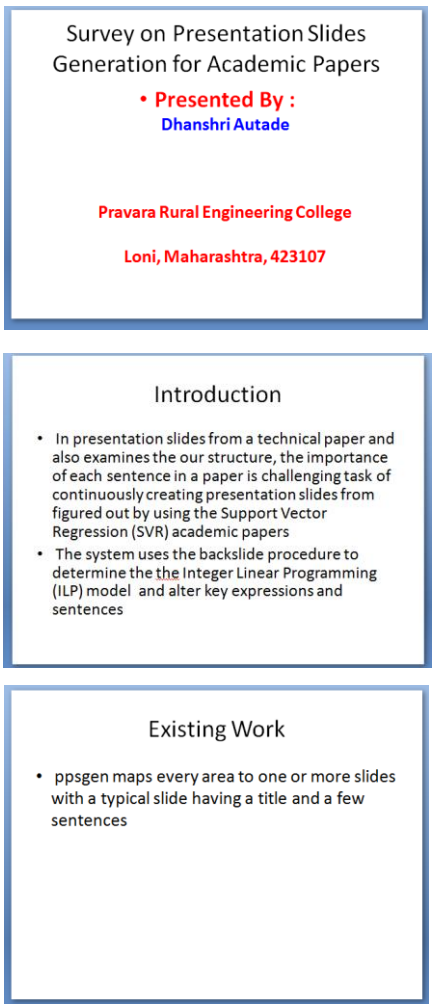
A. Comparison of Sentence Importance Computation Methods

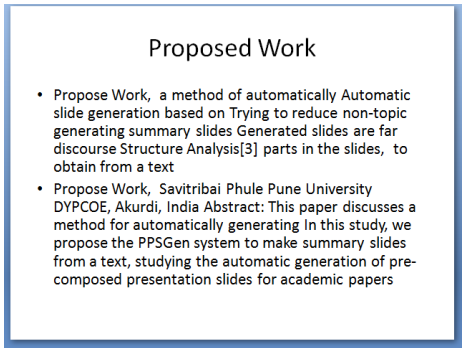
Table 1 shows the comparison results of the different sentence importance computation methods. Note that the sentence scores by the methods are as input of our ILP model for slides generation. We can see that the SVR-based methods (Overall-SVR, Avg-SVR, Our method) perform better than TF-IDF and Random Walk. It proves that the SVR model can better estimate the importance scores of the sentences. The SVR model is trained from large dataset and the sentences scores predicted by the SVR-based method can be more reliable to be used for slides generation. Among the three SVR-based methods, our method with the maximum similarity gets better ROUGE-1 and ROUGE- 2 values than those with the overall similarity or the average similarity. Generally, slides can be divided into several parts and each part may be relevant to one section in the paper. The sentences in a specific section should be more similar to the corresponding part in the slides and less similar to the other parts. Using the maximum similarity can reflect this intuition well. So it is better to use the maximum similarity as the sentence importance scoring method.

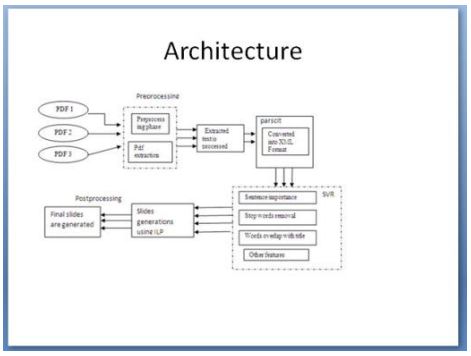
TABLE 1 ROUGE F-Measure Scores for Different ILP Models

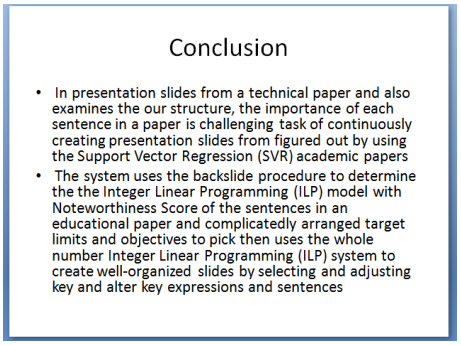
Method	Rouge-1	Rouge-2	Rouge-SU4
SILP1	0.40629	0.12599	0.17177
SILP2	0.41288	0.12906	0.17295
Our Method	0.41342	0.13067	0.17502

SAMPLE SLIDES









V. CONCLUSION

This paper proposes a novel framework called PPSGen to produce presentation slides from scholastic papers. Sentence scoring model is prepared taking into account SVR and use the ILP strategy to adjust and separate key expressions and sentences for producing the slides. Trial results demonstrate that our system can make immensely enhanced slides than standard schedules. scoring model taking into account SVR and utilize the ILP strategy to adjust and separate key expressions and sentences for creating the slides. Exploratory results demonstrate that our technique can create vastly improved slides than conventional strategie And also text and graphical elements in the paper and make slides more comprehensible and vivid also. In future work, we will improve our system by 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.

ACKNOWLEDGMENT

The authors would like to thank the Principal/ HOD/ Guide Colleague of the Department of Comp. Engg., P.R.E.C. COLLEGE, LONI Maharashtra, India, for their support, suggestions and encouragement.

REFERENCES

[1] Yue Hu and Xiaojun Wan PPSGen: Learning-Based Presentation Slides Generation for Academic Papers IEEE TRANSACTIONS ON NOWLEDGE AND DATA ENGINEERING,VOL. 27, NO. 4, APRIL 2015

[2] Y. Yasumura, M. Takeichi, and K. Nitta, A support system for making presentation slides, Trans. Japanese Soc. Artif. Intell., vol. 18, pp. 212220, 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. 754766.

[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. 102106.

[5] 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. 246249.

[6] 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. 240246.

[7] 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.

[8] M. Sravanthi, C. R. Chowdary, and P. S. Kumar, QueSTS: A query specific text summarization approach, in Proc. 21st Int. FLAIRS Conf., 2008, pp. 219224.

[9] N. Agarwal, K. Gvr, R. S. Reddy, and C. P. Rose, Towards multidocument summarization of scientific articles: Making interesting comparisons with SciSumm, in Proc. Workshop Autom. Summarization Different Genres, Media, Lang., 2011, pp. 815.

[10] O. Yeloglu, M. Evangelos, and Z.-H. Nur, Multi-document summarization of scientific corpora, in Proc. ACM Symp. Appl. Comput., 2011, pp. 252258.

[11] K. Woodsend and M. Lapata, Multiple aspect summarization using integer linear programming, in Proc. Joint Conf. Empirical Methods Nat. Lang. Process. 23 Comput. Nat. Lang. Learn., 2012, pp. 233243.

[12] D. Galanis, G. Lampouras, and I. Androutsopoulos, Extractive multi-document summarization with integer linear programming and support vector regression, in Proc.

[13] A. Clauset, M. E. Newman, and C. Moore, Finding community structure in very large networks. Phys. Rev. E, vol. 70, no. 6, p. 066111, 2004.

[14] G. Erkan and D. R. Radev, LexRank: Graph-based lexical centrality as salience in text summarization, J. Artif. Intell. Res., vol. 22,no. 1, pp. 457479, 2004.

[15] C. C. Chang and C. J. Lin. (2001), LIBSVM: A library for support vector machines, [Online]. Available: http: ((www.csie.ntu.edu. tw(cjlin(libsvm.

[16] A. Nenkova and R. J. Passonneau, Evaluating content selection in summarization: The pyramid method, in HLTNAACL, vol. 4, pp. 145152, May 2004

[17] S. Modgil, N. Faci, F. Meneguzzi, N. Oren, S. Miles, and M. Luck, A framework for monitoring agent-based normative systems, in Proc. 8th Int. Conf. Auton. Agents Multiagent Syst., 2009, pp. 153160

[18] V. Qazvinian, D. R. Radev, S. M. Mohammad, B. J. Dorr, D. M. Zajic, M. Whidby, and T. Moon, Generating extractive summaries of scientific paradigms, J. Artif. Intell. Res., vol. 46, pp. 165 201, 2013.

[19] 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. 555564.