Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

PPSGen: Learning to Generate Presentation Slides for Academic Papers

Yue Hu and Xiaojun Wan*
Institute of Computer Science and Technology
The MOE Key Laboratory of Computational Linguistics
Peking University
Beijing, China
{ayue.hu, wanxiaojun}@pku.edu.cn

Abstract

In this paper, we investigate a very challenging task of automatically generating presentation slides for academic papers. The generated presentation slides can be used as drafts to help the presenters prepare their formal slides in a quicker way. A novel system called PPSGen is proposed to address this task. It first employs regression methods to learn the importance of the sentences in an academic paper, and then exploits the integer linear programming (ILP) method to generate well-structured slides by selecting and aligning key phrases and sentences. Evaluation results on a test set of 200 pairs of papers and slides collected on the web demonstrate that our proposed PPSGen system can generate slides with better quality. A user study is also illustrated to show that PPSGen has a few evident advantages over baseline methods.

1 Introduction

Presentation slides have been a popular and effective means to present and transfer information, especially in academic conferences. The researchers always make use of slides to present their work in a pictorial way on the conferences. There are many softwares such as Microsoft PowerPoint and OpenOffice to help researchers prepare their slides. However, these tools only help them in the formatting of the slides, but not in the content. It still takes presenters much time to write the slides from scratch. In this work, we propose a method of automatically generating presentation slides for academic papers. We aim to automatically generate well-structured slides and provide such draft slides as a basis to reduce the presenter's time and effort when preparing their final presentation slides.

Academic papers always have a similar structure. They generally contain several sections like abstract, introduction, related work, proposed method, experiments and conclusions. 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. Our method attempts to generate draft slides of the typical type mentioned above and helps people to prepare their final slides.

Automatic slides generation for academic papers is a very challenging task. Current methods generally extract objects like sentences from the paper 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.

In this study, we propose a novel system called PPSGen 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.

Experiments on a test set of 200 paper-slides pairs indicate our method can generate slides with better quality than the baseline methods. Using the ROUGE toolkit, the slides generated by our method can get better ROUGE scores. Moreover, based on a user study, our slides can get higher rating scores by human judges in both content and structure aspects. Therefore, our slides are considered to be a better basis to prepare the final slides.

The rest of this paper is organized as follows. Related work is introduced in section 2. We describe our method in detail in section 3. We show the experiment results in section 4 and conclude our work in section 5.

2 Related Work

Automatic slides generation for academic papers remains far under-investigated nowadays. Few studies directly research on the topic of automatic slides generation. Masao *et al.* [1999] attempted to automatically generate slides from input

* Xiaojun Wan is the corresponding author.

documents annotated with the GDA tagset[1]. GDA tagging can be used to encode semantic structure. The semantic relations extracted between sentences include grammatical relations, thematic relations and rhetorical relations. They first detect topics in the input documents and then extract important sentences relevant to the topics to generate slides.

Yoshiaki *et al.* [2003] introduced a support system for making slides from technical papers. The inputs of the system are academic papers in LaTeX format. The system calculates the weights of the terms in the paper using TF*IDF scores. Using the term weights, objects in the paper like sentences, tables etc. are also weighted and used to determine the number of objects for each section to generate the slides. Shibata and Kurohashi [2005] proposed a method 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. These different parts are used to generate the final slides based on the detected discourse structure and some heuristic rules.

Hayama *et al.* [2005], Kan [2006] and Beamer and Girju [2009] studied the problem of aligning technical papers and presentation slides. Hayama *et al.* used a variation of the Hidden Markov Model (HMM) to align the text in the slides to the most likely section in the paper, which also used the additional information of titles and position gaps. Kan *et al.* applied a modified maximum similarity method to do the monotonic alignments and trained a classifier to detect slides which should not be aligned. Beamer and Girju compared and evaluated four different alignment methods that were combined by methods such as TF-IDF term weighting and query expansion.

Masum *et al.* [Masum *et al.*, 2005; Masum and Ishizuka, 2006] proposed a system named Automatic Report to Presentation (ARP) which constructed a topic-specific report and a presentation on a topic or search phrases given by a user.

Sravanthi *et al.* [2009] investigated automatic generation of presentation slides from technical papers in LaTeX. A query specific extractive summarizer QueSTS is used to extract sentences from the text in the paper to generate slides. QueSTS transfers the input text to an integrated graph (IG) where a sentence represents a node and edges exists between the nodes that the sentences corresponding to them are similar. More details can be found in [Sravanthi *et al.*, 2008].

Different from the above approaches which simply select and place objects like sentences on the slides, we attempt to generate more structured slides, which contain not only sentences but also key phrases aligned to the sentences. Key phrases are used as the bullet points and sentences relevant to the phrases are placed below them.

In addition, SVR and ILP have been used widely in the task of summarization. Ouyang *et al.* [2007] and Galanis *et al.* [2008] used SVR to learn the sentence score. McDonald [2007], Gillick *et al.* [2008 2009], Berg-Kirkpatrick *et al.* [2011] and Woodsend *et al.* [2012] adopted methods based on ILP to extract summary. Galanis *et al.* [2012] used a method based on both SVR and ILP to deal with multi-document summarization.

3 Our Proposed Method

3.1 Overview

In this paper, we propose a system to automatically generate slides that have good structure and content quality from academic papers. 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. More details of each part will be discussed in the following sections.

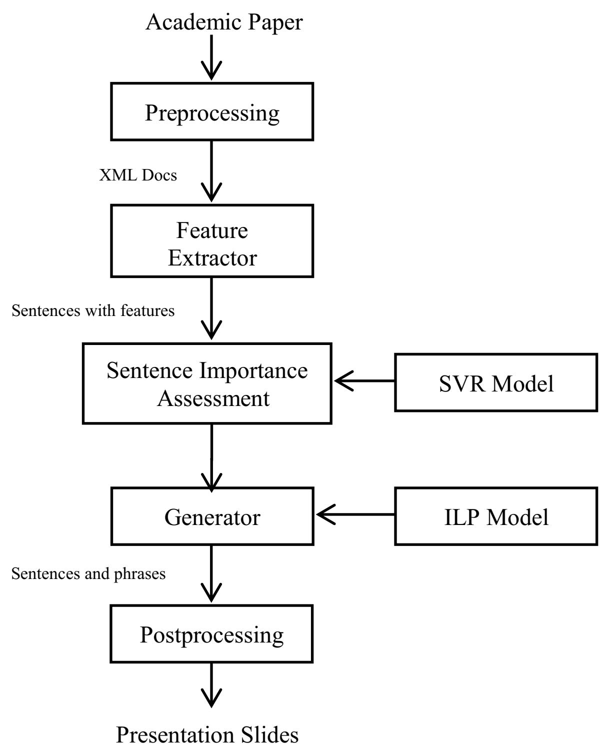

Figure 1: System Architecture

3.2 Corpus and Preprocessing

To learn how humans generate slides from academic papers, we build a corpus that contains pairs of academic papers and their corresponding slides. Many researchers in the computer science field place their papers and the corresponding slides together in their homepages. The homepages' URLs are obtained by crawling Arnetminer[2]. After downloading the homepages, we use several strict patterns to extract the links of the papers and the associated slides and download the files to build the dataset. We collect more than two thousand pairs.

[1] http://www.i-content.org/GDA/tagset.html

[2] http://arnetminer.org

After cleaning up the incorrect pairs, we have 1200 paper-slides pairs.

The papers are all in PDF format and the slides are in either PDF or PowerPoint format. For the papers, we extract their texts by using PDFlib[3] and detect their physical structures of paragraphs, subsections and sections by using ParsCit[4]. A custom XML format is used to describe this structure. For the slides, we also extract their texts and physical structures like sentences, titles, bullet points, etc. We use xpdf[5] and the API provided by Microsoft Office to deal with the slides in PDF and PowerPoint formats, respectively. The slides are transformed to a predefined XML format as well.

3.3 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.

Support Vector Regression Model

Here we briefly introduce the SVR model [Vapnik, 1998] . Let $\{u_i, y_i\}_{i=1}^{N}$ ($u_i \in R^d, y_i \in R$) be a set of inputs and outputs data points. The Support Vector Regression model aims to learn a function $f(u)$ which has the following form:

$$f(u) = w * \varphi(u) + b \quad (1)$$

where $\varphi(u)$ represents the high-dimensional feature spaces which are nonlinearly transformed from u. It chooses the optimum function $f(u)$ to minimizing the structure risk function below:

$$\frac{1}{2}\|w\|^2 + \frac{C}{N}\sum_{i=1}^{N} L_s(y_i, f(u_i)) \quad (2)$$

The first term is the regularized term to avoid overfitting. The other term is the empirical error measured by ε-insensitivity loss function which is defined as $L(x) = |x| - \varepsilon$ if $x > \varepsilon$, and $L(x) = 0$, otherwise. The regularization constant C and the radius ε can be set by the user.

After introducing the kernel function $k(u_i, u_j)$ and solving the optimization problem mentioned above, we can get

$$f(u) = \sum_{i=1}^{n} \beta_i k(u_i, u) + b \quad (3)$$

In our case, u is the feature vector of the sentence and y is the importance score of the sentence. We use LIBSVM [Chang and Lin, 2001] with a RBF kernel to implement the SVR model.

Training Data Construction and Model Learning

To construct training data based on the paper-slides pairs, we apply a similarity scoring method to assign the importance scores to the sentences in a paper. The main hypothesis is that the sentences in the slides should represent the substance of the corresponding paper. The sentences in the paper which are more similar to the sentences in the slides should be considered more important and higher scores should be assigned to them using the scoring method.

Thus, we define one sentence's importance score in the paper as follows:

$$score(s) = max_{s_i^* \in S^*}(sim(s, s_i^*)) \quad (4)$$

where s is a sentence in the paper, S^* is the set of the sentences in the corresponding slides. The standard cosine measure is used as the similarity function $sim(s, s_i^*)$.

We adopt the maximum similarity instead of the average similarity with all the sentences in the slides. The motivation is that slides generally 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. Therefore, it is more reasonable to use the maximum similarity to assign the importance scores of the sentences.

Each sentence is represented by a set of features. In this study, we make use of the following features of each sentence s:

1) Similarity with the titles:

We consider three types of titles: paper title, section titles and subsection titles. Only the titles of the section and subsection which contains the sentence are used. We use the cosine similarity values between the sentence and different types of titles as different features. Stop words are removed and all the words are stemmed in the similarity calculation.

2) Word overlap with the titles:

It is the number of words shared by the sentence and the set of all titles, including all three types of titles.

3) Sentence position: It is computed as follows:

$$SP(s) = \frac{pos(s, sec(s))}{|sec(s)|}$$

where $pos(s, sec(s))$ is the position of sentence s in its section, $|sec(s)|$ is the number of sentences in $sec(s)$.

4) Sentence's parser tree information:

The features are extracted from the sentence's parse tree. It includes the number of noun phrases and verb phrases, the number of sub-sentences and the depth of the parse tree.

5) Stop words percentage:

It is the percentage of the stop words in the total word set of the sentence s.

6) Other features including the length of sentence s and the number of words after removing stop words.

All the features mentioned above are scaled into [-1, 1]. Based on the features and importance scores of the sentences in the training data, we can learn a SVR model, and then apply the model to predict an importance score for each sentence in any paper in the test set. The score indicates the possibility of a sentence to be selected for making slides.

[3] http://www.pdflib.com/
[4] http://aye.comp.nus.edu.sg/parsCit/
[5] http://www.foolabs.com/xpdf/

3.4 Slides Generation

After getting the predicted importance score for each sentence in the given paper, we exploit the Integer Linear Programming (ILP) method to generate well-structured slides by selecting and aligning key phrases and sentences.

Unlike those methods [Masao *et al.*, 1999; Yoshiaki *et al.*, 2003; Sravanthi *et al.*, 2009] that generate slides by simply selecting important sentences and placing sentences on the slides, we select both key phrases and sentences to construct well-structured slides. We use key phrases as the bullet points and sentences relevant to the phrases are placed below the bullet points.

In order to extract the key phrases, chunking implemented by the OpenNLP[6] library is applied to the sentences and noun phrases are extracted as the candidate key phrases.

We define two kinds of phrases: global phrases and local phrases. A local phrase means a global phrase in a particular section. A unique phrase (e.g. "SVR") has a global phrase identifier ("SVR"), while its appearances in different sections are considered as different local phrases (e.g. "SVR_introduction", "SVR_related_work"). A global phrase that corresponds to more local phrases should be regarded to be more important and more likely to be selected. Thus, we use the local phrases to generate the bullet points directly for different sections and use the global phrases to address the importance differences between different unique phrases. All the phrases are stemmed and stop words are removed. Moreover, the noun phrases that appear only once in the paper are discarded.

The object function and constraints of our proposed ILP Model are presented as follows:

$$\max_{lp,x} \lambda_1 \sum_{i=1}^{n} \frac{l_i}{L_{max}} w_i x_i + \lambda_2 \sum_{i=1}^{|B|} \frac{c_{b_i} b_i}{|B^*|} + \lambda_3 \sum_{i=1}^{n} \frac{w_i}{n} y_i \quad (5)$$

Subject to:

$$\sum_{i=1}^{n} l_i x_i \le L_{max} \quad (6)$$

$$\sum_{lp_j \in LP_i} lp_j \ge x_i, \text{ for } i = 1,\dots n \quad (7)$$

$$\sum_{s_i \in S_j} x_i \ge lp_j, \text{ for } j = 1,\dots |LP| \quad (8)$$

$$\sum_{lp_j \in LP_k} lp_j \ge y_k, \text{ for } k = 1,\dots n \quad (9)$$

$$\sum_{b_m \in B_i} b_m \ge |B_i| x_i, \text{ for } i = 1,\dots n \quad (10)$$

$$\sum_{s_i \in S_m} x_i \ge b_m, \text{ for } m = 1,\dots |B| \quad (11)$$

$$\sum_{lp_j \in gp_t} lp_j \ge gp_t, \text{ for } t = 1,\dots |GP| \quad (12)$$

$$gp_t \ge lp_j, \text{ for } \forall lp_j \in gp_t; \ t = 1,\dots |GP| \quad (13)$$

$$\sum_{i=1}^{|GP|} gp_i * 2 \le \sum_{i=1}^{n} x_i \quad (14)$$

$$x_i, lp_j, y_k, b_m, gp_t \in \{0,1\}, \ \forall i,j,k,m \quad (15)$$

Where:

w_i	-	the importance weight of sentence s_i which is computed by using the SVR model;
n	-	the number of sentences
l_i	-	the length of sentence s_i
x_i	-	the variable that indicates whether sentence s_i is included in the slides
lp_j	-	the variable that indicates whether local phrase lp_j is included in the slides
gp_t	-	the variable that indicates whether global phrase gp_t is included in the slides
y_k	-	the variable that indicates whether sentence s_k contains at least one selected local phrase
b_m	-	the variable that indicates whether bigram b_m is included in the slides
L_{max}	-	the maximum word count of the slides
c_{b_i}	-	the count of the occurrences of bigram b_i in the paper
B^*	-	the total set of bigrams in the paper
B	-	the set of unique bigrams after removing duplicated bigrams in B^*
LP	-	the set of the total local phrases
GP	-	the set of the total global phrases

The object function contains three parts. The explanation of each part is below:

- The first part maximizes the overall importance score of the generated slides. It sums the importance scores of the selected sentences. Rather than simply calculating the sum of the scores, we add the sentence length as a multiplication factor in order to penalize the very short sentences.
- The second part maximizes the total weights of the bigrams in the paper which also appear in the slides. The intuition is that when more bigrams are present in the slides, the sentences in the slides are less redundant. c_{b_i} can be regarded as the weight of the bigram. We try to include more important bigrams in the slides.
- The last part aims to maximize the weighted coverage of the key phrases selected. A sentence is covered by a phrase when this sentence contains the phrase. High-quality slides should cover the content in the paper as much as possible. We describe this kind of coverage by using the sum of the scores of the sentences that contains the selected key phrases.

All the terms in the object function are normalized to [0, 1] by using the maximum length L_{max}, the total number of bigrams $|B^*|$ and the number of the sentences n, respectively. The values of λ_1, λ_2 and λ_3 are parameters for tuning the three parts and we set $\lambda_1+\lambda_2+\lambda_3 = 1$.

The explanation of each constraint is below:

Constraint (6): It guarantees that the total word count of the slides does not exceed L_{max}.

Constraint (7): LP_i is the set of phrases that sentence s_i contains. It ensures that when sentence s_i is selected, at least one local phrase in LP_i is selected.

[6] http://opennlp.apache.org/

Constraint (8): S_j is the set of sentences that contains phrase lp_j. It ensures that when local phrase lp_j is selected, at least one sentence in S_j is selected.

Constraint (9): LP_k is the set of phrases that sentence s_k contains. It ensures that y_k is set to 1 only when at least one local phrase in LP_k is selected.

Constraints (10), (11): These two constraints are similar to constraints (8) and (9), respectively. B_i is the set of bigrams that sentence s_i contains. S_m is the set of sentences that include bigram b_m. When constraint (10) holds, the total bigrams that s_i has are selected if s_i is selected. Constraint (11) guarantees at least one sentence in S_m is selected if b_m is selected.

Constraints (12), (13): These two constraints ensure that if a local phrase is selected, the corresponding global phrase is also selected. Meanwhile, if a global phrase is selected, at least one corresponding local phrase is selected.

Constraints (14): It guarantees the total number of the selected global phrases is less than half the number of the sentences selected. Using this constraint, we can avoid extracting too many key phrases. We use the count of global phrases instead of local phrases. The motivation is that phrases appear in several sections are considered more important and we reward these phrases by this way.

The ILP model is applied to the whole paper once and the method does not assign the number of slides for each section explicitly. Using the ILP model, we can obtain the aligned key phrases and sentences to be included in the slides. The titles of slides are set by using the titles of the corresponding sections. We solve the above optimization problem by using the IBM CPLEX optimizer [7]. It generally takes about ten seconds to solve the problem. Then the draft slides are generated by using the API provided by Microsoft Office.

4 Evaluation

4.1 Evaluation Setup

In order to set up our experiments, we divide our dataset which contains 1200 pairs of paper and slides into two parts: 1000 pairs for training and the other 200 pairs for test. A SVR regression model with the RBF kernel in LIBSVM is trained on the training data and applied to the test data. Then the ILP model is used to generate the slides. The maximum word count of the slides is simply set to fifteen percentage of the total word count of the paper. The parameter values of our method are empirically set to 0.3, 0.4 and 0.3, respectively.

We implement three baseline methods for comparison with our proposed method. The first method is the one used by [Yoshiaki *et al.*, 2003] based upon the TF-IDF scores, which extracts sentences for each section or subsection. The IDF scores are calculated on the corpus we collect. The other two methods are based on traditional summarization methods: MEAD[8] and Random Walk. MEAD [Radev *et al.*, 2004] is the most elaborate and publicly available platform for multi-lingual summarization. It implements multiple summarization methods including position-based, centroid-based, length-based and query-based. A combination of these methods is adopted to extract the slides. In the Random Walk [Page *et al.*, 1998] method, sentences are regarded as nodes, the cosine similarities between sentences are assigned to be the weights of edges and the random walk method is employed to extract sentences for each section or subsection. All the slides that extracted by these baseline methods have the same lengths as those generated by our method. The total length of sentences extracted for each section is determined by the section's length to the paper's length radio in the TF-IDF based and Random Walk baseline methods.

First, we use the ROUGE toolkit to evaluate the content quality of the generated slides. ROUGE [Lin, 2004] is a state-of-the-art automatic evaluation method based upon *n*-gram comparison, which is widely used in summarization evaluation. Here, the set of sentences in the author-written slides of the paper is regarded as the model text. We use the F-Measure score of ROUGE-1, ROUGE-2, ROUGE-SU4. ROUGE-1 and ROUGE-2 evaluates a system generated text by matching its unigrams and bigrams against the model texts, respectively. ROUGE-SU4 matches unigrams and skip-bigrams of a generated text against model texts. Skip-bigram is a pair of words in the sentence order, allowing for gaps within a limited size which is always set to four. When evaluating the slides, stop words are removed and stemming is utilized. We use two different ways to evaluate the slides. The first way is to evaluate and compare the whole generated slides and author-written slides, while the second one is to evaluate and compare the first 30 percentage, the middle 40 percentage and the last 30 percentage of both types of slides, respectively. We aim to better evaluate the coverage and structure of the generated slides by doing that.

In addition, paired T-Tests are applied between the ROUGE scores obtained by each baseline and our method in the first evaluation way .

Moreover, a user study is also performed to subjectively evaluate the slides generated by different methods. We randomly select 20 papers in the test set and employ Yoshiaki's method, MEAD and our system to generate slides. Random Walk is skipped because it is also a summarization method and gets similar performance as MEAD. These slides are presented to four human judges. The judges are asked to answer the following questions by giving a rating on a scale of 1 to 5 for the presentations (5 means very good, 1 means very bad):

1. What is your satisfaction level of the slides' structure?
2. What is your satisfaction level of the slides' content?
3. What is your overall satisfaction level on the slides?

4.2 Results and Discussion

The comparison results over ROUGE metrics are presented in Table 1 and Table 2. Table 1 shows that our proposed method can get better ROUGE scores, i.e., better content

[7] http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/

[8] http://www.summarization.com/mead/

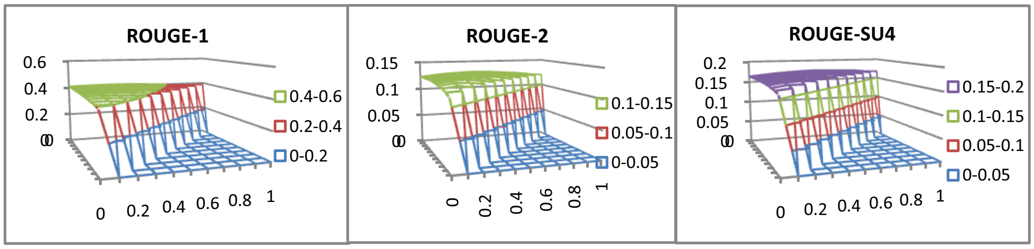

Figure 2: Parameter influences (horizontal, vertical axis are λ_1, λ_2 , respectively, $\lambda_3 = 1 - \lambda_1 - \lambda_2$)

Method	Rouge-1	Rouge-2	Rouge-SU4
Yoshiaki *et al.*	0.38859	0.11624	0.16424
Random Walk	0.39421	0.11555	0.16463
Mead	0.38778	0.11803	0.16239
Our Method	**0.41342**	**0.13067**	**0.17502**

Table 1: ROUGE F-measure scores obtained in the first way

Method	Rouge 1			
	First 30%	Mid 40%	Last 30%	Avg
Yoshiaki *et al.*	0.28220	0.30732	0.28411	0.29121
Random Walk	0.29241	0.30661	0.28443	0.29448
Mead	0.31132	0.28063	0.25481	0.28225
Our Method	**0.30235**	**0.32662**	**0.29911**	**0.30936**

Table 2: ROUGE-1 F-measure scores obtained in the second way

System (tested by our method)	T-Test		
	Rouge-1	Rouge-2	Rouge-SU4
Yoshiaki *et al.*	2.492E-13	2.749E-08	1.015E-06
Random Walk	5.329E-06	6.976E-11	1.125E-06
Mead	2.176E-11	1.601E-11	6.331E-08

Table 3: T-Test p-values between each baseline method and our method

Judge	Method	Structure	Content	Overall
avg	Yoshiaki *et al.*	3	2.68	2.71
	Mead	2.7	2.53	2.39
	Our Method	**3.69**	**3.45**	**3.58**

Table 4: Average rating scores of judges

quality in the first evaluation way. It means that our slides are richer in content and much more similar to the human-written slides than those of the baselines as a whole. Table 2 shows that our method can mainly get better ROUGE scores in each part comparison. It means that our slides can also achieve better content quality when they are divided into continuous parts and corresponding parts are compared, i.e., the content texts in our method are more effectively distributed into different sections. We obtain the importance score of a sentence by learning from the human slides. Therefore, the importance scores are more credible. Moreover, the alignment between key phrases and sentences is also useful to improve the content, because the sentences relevant to such key phrases can be considered to be more important. We distinguish global phrase from local phrase. Sentences that contain more important global phrases can be regard to be more important. Using global phrase can lead to a better selection of key sentences, which also result in better content quality.

Table 3 proves our method's improvements are statistically significant, and the T-Test values we get are all far smaller than 0.05. Figure 2 presents the influences of ROUGE scores when tuning the parameters λ_1, λ_2 and λ_3. We set the sum of the three parameters to one, and thus we actually need to change two of the three parameters. We can see that when the parameters are set in a wide range of values, our method can achieve high ROUGE scores. Our system generally performs better than those baselines. We can also see that all the three parts in our ILP model are helpful to get a better content quality for the generated slides. Without any part of them, the results will get worse.

Table 4 shows the average scores rated by human judgers for each method. The slides generated by our method obviously have better overall quality than those of other methods. Being consistent with the automatic evaluation results, our slides are considered to have better content quality according to human judges. Moreover, owing to the indent structure and the alignment between phrases and sentences, the structure of our slides is also judged to be much better than the baselines' slides.

Overall, the experimental results indicate that our method can generate much better slides than the baselines in both automatic and human evaluations.

5 Conclusions and Future Work

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 extract aligned key phrases and sentences for generating the slides. Experimental results show that our method can generate much better slides than traditional methods.

In the future work, we will improve our system by using both text and graphical elements in the paper and make slides more comprehensible and vivid.

Acknowledgments

The work was supported by NSFC (61170166), Beijing Nova Program (2008B03) and National High-Tech R&D Program (2012AA011101).

References

[Beamer and Girju, 2009] Brandon Beamer and Roxana Girju. Investigating Automatic Alignment Methods for Slide Generation from Academic Papers. Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL), pages 111–119, Boulder, Colorado, June 2009.

[Berg-Kirkpatrick *et al.*, 2011] Berg-Kirkpatrick, T., Gillick, D., and Klein, D. Jointly learning to extract and compress. In Proceedings of the 49th Annual Meeting of the ACL-HLT, pages 481–490, Portland, OR.

[Chang and Lin, 2001] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A Library for Support Vector Machines, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm

[Galanis *et al.*, 2008] Galanis, D. and Malakasiotis, P. AUEB at TAC 2008. In Proceedings of the Text Analysis Conference, Gaithersburg, MD.

[Galanis *et al.*, 2012] Dimitrios Galanis, Gerasimos Lampouras and Ion Androutsopoulos. Extractive Multi-Document Summarization with Integer Linear Programming and Support Vector Regression.

[Gillick *et al.*, 2008] Gillick, D., Favre, B., and Hakkani-Tur, D. The ICSI Summarization System at TAC 2008. In *Proceedings of the Text Analysis Conference*, Gaithersburg, MD.

[Gillick *et al.*, 2009] Gillick, D. and Favre, B. A scalable global model for summarization. In Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, pages 10–18, Boulder, CO.

[Hayama *et al.*, 2005] Hayama, T., Nanba, H., Kunifuji, S. Alignment between a technical paper and presentation sheets using hidden markov model. In: Proceedings of the 2005 International Conference on Active Media Technology.

[Kan, 2006] Min-Yen Kan. SlideSeer: A digital library of aligned document and presentation pairs. JCDL'06, June 11–15, 2006, Chapel Hill, North Carolina, USA.

[Lin, 2004] Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL 2004, Barcelona, Spain.

[Masao *et al.*, 1999] Masao, U., and Koiti, H. Automatic slide presentation from semantically annotated documents. In Proceedings of the ACL workshop on Coreference and its applications. Morristown, NJ, USA: Association for Computational Linguistics.

[Masum *et al.*, 2005] Masum, S. M. A.; Ishizuka, M.; and Islam, M. T. Auto-presentation: A multi-agent system for building automatic multi-modal presentation of a topic from world wide web information. Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology.

[Masum and Ishizuka, 2006] Masum, S. M. A., and Ishizuka, M. Making topic specific report and multimodal presentation automatically by mining the web resources. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence.

[McDonald, 2007] McDonald, R. A study of global inference algorithms in multi-document summarization.In Proceedings of the European Conference on Information Retrieval, pages 557–564, Rome, Italy.

[Modgil *et al.*, 2009] Sanjay Modgil, Noura Faci, Felipe Rech Meneguzzi, Nir Oren, Simon Miles, and Michael Luck. A framework for monitoring agent-based normative systems.

[Ouyang *et al.*, 2007] You Ouyang, Sujian Li and Wenjie Li. Developing Learning Strategies for Topic-based Summarization. CIKM'07, November 6–8, 2007, Lisboa, Portugal.

[Page *et al.*, 1998] L. Page, S. Brin, R. Motwani and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Libraries, 1998.

[Radev *et al.*, 2004] Dragomir R. Radev, Timothy Allison, et al. MEAD - a platform for multidocument multilingual text summarization. In Proceedings of 4th International Conference on Language Resources and Evaluation.

[Shibata and Kurohashi, 2005] Shibata, T., and Kurohashi, S. Automatic slide generation based on discourse structure analysis. In IJCNLP, 754–766.

[Sravanthi *et al.*, 2009] M. Sravanthi, C. Ravindranath Chowdary and P. Sreenivasa Kumar. SlidesGen: Automatic Generation of Presentation Slides for a Technical Paper Using Summarization. Proceedings of the Twenty-Second International FLAIRS Conference (2009).

[Sravanthi *et al.*, 2008] Sravanthi, M.; Chowdary, C. R.; and Kumar, P. S. QueSTS: A query specific text summarization approach. In Proceedings of the 21st International FLAIRS Conference, 219–224. Florida, USA: AAAI Press.

[Vapnik, 1998] Vapnik, V. (1998). Statistical Learning Theory. John Wiley.

[Woodsend *et al.*, 2012] Woodsend, K. and Lapata, M. Multiple aspect summarization using integer linear programming. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 233–243, Jesu Island, Korea.

[Yoshiaki *et al.*, 2003] Yoshiaki, Y., Masashi, T., Katsumi, N. A support system for making presentation slides (in Japanese). Transactions of the Japanese Society for Artificial Intelligence 18 (2003) 212-220