Keyphrase Extraction from Online News Using Binary Integer Programming

Zhuoye Ding, Qi Zhang, Xuanjing Huang
Fudan University
School of Computer Science
{09110240024,qz,xjhuang}@fudan.edu.cn

Abstract

In recent years, keyphrase extraction has received great attention, and been successfully employed by various applications. Keyphrases extracted from news articles can be used to concisely represent main contents of news events. Keyphrases can help users to speed up browsing and find the desired contents more quickly. In this paper, we first present several criteria of high-quality news keyphrases. After that, in order to integrate those criteria into the keyphrase extraction task, we propose a novel formulation which converts the task to a binary integer programming problem. The formulation cannot only encode the prior knowledge as constraints, but also learn constraints from data. We evaluate the proposed approach on a manually labeled corpus. Experimental results demonstrate that our approach achieves better performances compared with the state-of-the-art methods.

1 Introduction

Keyphrase extraction is a long studied topic in natural language processing. A keyphrase, which consists a word or a group of words, is defined as a precise and concise expression of one or more documents. It has been widely used in various applications such as summarization, clustering, categorizing, browsing, and so on. In recent years, keyphrase extraction has received much attention (Witten et al., 1999; Zha, 2002; Hulth, 2003; Tomokiyo and Hurst, 2003; Chen et al., 2005; Medelyan et al., 2009; Liu et al., 2009).

Keyphrases are usually manually chosen by authors, for scientific publications, magazine articles, books, et al. Due to the expensive and time consuming effort of manually assigning keyphrase, web pages and online news rarely contain keyphrases. It should be useful to automatically extract keyphrases from online news to represent their main contents. There are already a number of studies which focus on extracting keyphrases from scientific publications or single news article (Frank et al., 1999; Turney, 2000; Wan and Xiao, 2008; Jiang et al., 2009). We also notice that, currently, many websites provide the service which group related news together to facilitate users' browsing. In this paper, we focus on extracting keyphrases from a group of news articles which describe the same news event by different publishers.

Previous studies on keyphrase extraction can be roughly categorized into two groups: supervised and unsupervised. Unsupervised approaches usually select a set of candidates and use different ranking methods to select the candidates with the highest scores as keyphrases. Most of ranking methods are based on the information extracted from the document, such as TF·IDF, position, syntactic relation with other words, and so on. Supervised methods convert the task into a binary classification problem, which categorizes phrases as keyphrases or non-keyphrases. Similar as other tasks applied by supervised methods, a large amount of domain dependent training data is required. When the domain is changed, the labeled corpus should also be changed. And corpus labeling is a time-consuming and tedious task.

Most of the current methods focus on judging the importance of each phrase, and individually extract phrases with the highest scores. After analyzing the human assigned keyphrases, we observe

Proceedings of the 5th International Joint Conference on Natural Language Processing, pages 165–173,
Chiang Mai, Thailand, November 8 – 13, 2011. ©2011 AFNLP

that the keyphrases of news should satisfy the following properties:

1. **Relevance**. The keyphrases should be semantically relevant to the news theme. The most important ones should be selected as keyphrases.
2. **Coverage**. The keyphrases should be indicative of the whole news event. The extracted keyphrases should cover most of the aspects of the news event.
3. **Coherence**. The keyphrases should be semantically related to each other, and logically consistent and holding together as a harmonious whole.
4. **Conciseness**. The keyphrases should not contain keyphrases with redundant information.

In order to automatically select keyphrases which can satisfy the above properties, in this paper, we propose a novel formulation which converts keyphrase extraction to a binary integer programming problem (BIP) (Alevras and Padberg, 2001). An objective function and a number of constraints which high-quality keyphrases should satisfy are specified. BIP, which is the special case of integer programming and a well-studied optimization framework, is used to efficiently search the entire space to extract keyphrases. The formulation provides a flexible framework for integrating different criteria as objective functions or constraints.

The major contributions of this work can be summarized as follows: 1) We propose a novel formulation of keyphrase extraction as a binary integer programming problem; 2) Several criteria which high-quality keyphrases should satisfy are converted to the objective function and a set of constraints in order to fit the formulation; 3) Keyphrases are extracted as a set with consideration of their relationships; 4) Experimental results on the dataset consisting of 150 groups of news articles with human annotated keyphrases demonstrate that the proposed method performs better than the state-of-the-art algorithms.

The rest of this paper is organized as follows: Section 2 reviews some related studies. We propose our approach in Section 3. In Section 4, the experimental results are shown and discussed. Finally, we conclude this paper in Section 5.

2 Related Work

As mentioned in the previous section, most of current studies on keyphrase extraction can be roughly divided into two categories: supervised and unsupervised approaches.

Unsupervised approaches usually select general sets of candidates and use a ranking step to select the most important candidates. For example, Mihalcea and Tarau proposed a graph-based approach called TextRank, where the graph nodes are tokens and the edges reflect cooccurrence relations between tokens in the document (Mihalcea and Tarau, 2004). Wan and Xiao expanded TextRank by using a small number of topic-related documents to provide more knowledge, which improved results compared with standard TextRank and a tf.idf baseline (Wan and Xiao, 2008). Tomokiyo and Hurst used pointwise KL-divergence between language models derived from the documents and a reference corpus (Tomokiyo and Hurst, 2003). Matsuo and Ishizuka presented a statistical keyphrases extraction approach that did not make use of a reference corpus, but was based on cooccurrences of terms in a single document (Y.Matsuo and M.Ishizuka, 2004). In this paper the proposed BIP based method can combine those unsupervised methods as assignment value in the objective function. TF·IDF and locality information are used in our approach.

Supervised approaches use a corpus of training data to learn a keyphrase extraction model that is able to classify candidates as keyphrases or non-keyphrases. A well known supervised system is KEA that uses all n-grams of a certain length as candidates, and ranks them based on a Naive Bayes classifier using tf.idf and position as its features (Frank et al., 1999). Then Medelyan and Witten presented the improved KEA++ that selected candidates with reference to a controlled vocabulary from a thesaurus or Wikipedia (Medelyan and Witten, 2006). “Extractor” was another supervised system that used stems and stemmed n-grams as candidates (Turney, 2000). Its features are tuned using a genetic algorithm. Turney introduced a feature set based on statistical word association to ensure that the returned keyphrases set is coherent (Turney, 2003). Experimental results showed that coherence features can significantly improve the performance and they were not domain-specific. Nguyen and Kan presented a keyphrase extraction algorithm for scientific publications and introduced novel features towards scientific publications such as section information and certain morphological phenomena often found in scientific papers (T.D.Nguyen and Kan., 2007).

Since integer linear programming (Alevras and Padberg, 2001) can be used to incorporate both local features and non-local features, which are difficult to handle with traditional algorithms, it has received much attention in various NLP problems in recent years. Roth and Yih (2005) extended CRF models by applying inference procedure based on ILP to naturally and efficiently support general constraint structures. They applied their model on semantic role labeling (SRL) task. Martin et al. (2009) formulated the problem of nonprojective dependency parsing as a polynomial-sized integer linear program. Woodsend and Lapata (2010) presented a joint content selection and compression model for single-document summarization using an integer linear programming formulation.

3 Keyphrase Extraction Using BIP

The objective of keyphrase extraction is to select the most informative group of phrases, which are relevant to the news event and subject to constraints including the number of phrases, topic/aspect coverage, and coherence. Since these constraints are global, and cannot be adequately satisfied by optimizing each of them individually, our approach uses the BIP formulation, a well-studied optimization framework, which can be efficiently solved using standard optimization tools, to extract keyphrases.

Integer Linear Programming (ILP) denotes a set of constraint optimization problems which have a linear objective function, subject to linear equality and linear inequality constraints, and require the objective variables to be integers. ILP can be expressed in canonical form:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \le b \\ & Gx = d \\ & x \in \mathbf{Z}^n \end{aligned} \tag{1}$$

Binary Integer Programming (BIP) is the special case of ILP where variables are either 0 or 1.

In this paper, we treat the keyphrase extraction task as a two class labeling problem. Given a group of documents D, for each word $w \in D$, we decide to select this word as a keyphrase (assign label "1" to the word), or non-keyphrase (assign label "0"). We use a vector of binary variables $x = (x_1, x_2, ..., x_n)$ over word $w_i \in D$, to indicate whether the corresponding word should be selected or not. With the objective variables x and word $w_i \in D$, $c = (c_1, c_2, ..., c_n)$ is defined as the assignment value. The variable c_i gives the expected value of labeling w_i as a keyphrase. The basic extraction model is shown in Eq.(2). Our goal is to find the optimal point of weights x^* satisfying the constraints.

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & 0 \le x_i \le 1 \\ & x \in \mathbf{Z}^n \end{aligned} \tag{2}$$

3.1 Objective Function

With the BIP formulation, objective function $c^T x = \sum_k c_k x_k$ denotes the expected informative scores over all the words of a solution x. Maximizing the expected scores biases the words with highest c_i values as keyphrases. Various features can be considered as the values c. In this work, we use two basic features TF·IDF and locality. They have also been widely used in existing keyphrase extraction methods. The objective function is given in the Eq.(3).

$$\begin{aligned} c^T x, \ \ c_i = \quad & \alpha \cdot \frac{\sum_{d \in D} TF{\cdot}IDF(w_i, d)}{|D|} \\ & + \beta \cdot \mu_i + \gamma \cdot \nu_i \end{aligned} \tag{3}$$

Three parameters α,β, and γ are used to tradeoff among the different parts, $|D|$ is the number of documents in this news group. The latter section provides detailed description of this equation.

3.1.1 TF·IDF

TF·IDF compares the frequency of a phrase in a particular document with that in general corpus. The TF·IDF for word w_i is computed as:

$$TF{\cdot}IDF(w_i, d) = \frac{\text{freq}(w_i, d)}{|d|} \cdot \log_2 \frac{N}{df(w_i)}, \text{ where}$$

freq(w_i, d) is the number of times w_i occurs in d; $df(w_i)$ is the number of documents containing w_i in the global corpus; N is the size of the global corpus; $|d|$ is the length of the document of d..

In this paper, we use the average TF·IDF over all the news articles belonging to the same group. TF·IDF has also been used as features by almost all the keyphrase extraction algorithms.

3.1.2 Locality

The first occurrence position of the candidate phrase is an important feature for keyphrase extraction. It has also been used by many existing methods (Witten et al., 1999; Zha, 2002; Liu et al., 2009). In this paper, we also incorporate the information as parts of objective function.

For the words in the title of news articles, we define a bonus μ for their informative scores. It is the second component in the Eq.(3). The μ_i is defined as follows:

$$\mu_i = \begin{cases} \mu, & w_i \in T \\ 0, & \text{otherwise} \end{cases}$$

, where T represents the set of all the title words.

Similarly, we define ν for those words which occur in the first sentences. It is the third component of the objective function. The ν_i is defined as follows:

$$\nu_i = \begin{cases} \nu, & w_i \in FS \\ 0, & \text{otherwise} \end{cases}$$

,where FS represents the set of words which occur in the first sentences.

3.2 Constraints

One limitation of existing keyphrase extraction methods is that they usually separately make judgment of individual phrase instead of considering the qualities of the set of phrases as a whole. In this section, we define several constraints converted from the coverage and coherence criteria, and the number of extracted phrases.

3.2.1 Coverage

From both observations we make, and the properties proposed by Liu et al.(2009), we believe that high-quality keyphrases should cover the whole document or group of documents well. For example, if we have a document describing “Toyota recalls Prius” from various aspects of “reason”, “scope”, “influence” and so on., the extracted keyphrases should cover as many aspects as possible.

In order to satisfy this criterion, topic model is used to estimate words distribution over topics. In this paper, we use latent Dirichlet allocation (LDA) (Blei et al., 2003) to do it[1]. LDA is a three-level hierarchical Bayesian model, in which each word is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.

From LDA model, we can get $p(w|z)$, which represents aspect distributions over words. It indicates which words are important to an aspect. We use matrix G to represent $p(w|z)$. The vector g_i denote the distribution over words of aspect i. The projection $g_i^T x$ gives us the aspect coverage of topic i under current solution x. We want the coverage of every aspect to exceed the same threshold ζ. The constraint can be expressed as follows:

$$G^T x \succeq \zeta$$

3.2.2 Coherence

According to the properties which high-quality keyphrases should satisfy, the keyphrases should be semantically related and coherent. Turney (2003) also mentioned this issue and pointed out that incoherent keyphrases might highly impact the quality and user experience.

An intuitive method for measuring word relations is based on word cooccurrence relations within the document. It indicates that word pairs with high cooccurrence frequency should be selected together. For instance, the words “economy”, “unemployment”, and “loan” are likely to cooccur in documents about “financial crisis”. And we are aiming to extract them together to ensure coherence property. In this paper, we use mutual information (MI) to measure the word’s coherence. MI is a measure of association which quantifies the discrepancy between the dependent joint distribution and the independent individual distributions.

For each word pair $< w_i, w_j >$, whose mutual information $I(w_i, w_j)$ is bigger than a pre-defined threshold ξ, we add the following constraint:

$$x_i - x_j = 0$$

It encodes the fact that keyphrases pairs with high cooccurrence frequency should be selected together.

3.2.3 Number of Extracted Phrases

According to the limitations of space or other constraints given by applications, the number of extracted phrases should also be constrained. Since we use a vector of binary variables $x = (x_1, x_2, ..., x_n)$ over words $w_i \in D$, the constraint

[1] We use MALLET 2.0.6 in the experiments

can be represented as follows:

$$\sum_{i=1}^{n} x_i \leq K$$

,where K is the pre-defined threshold.

3.3 BIP Problem

Putting the objective function and all the constraints together, we obtain the BIP program to extract keyphrases as follows:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & G^T x \succeq \zeta \\ & x_i - x_j = 0\ , \ \text{ if } \ I(w_i, w_j) \geq \xi \\ & \sum_{i=1}^{n} x_i \leq K \qquad\qquad (4) \\ & x_i \in \{0,1\}\ , \ \ i = 1 \cdots n \end{aligned}$$

Binary integer programming is a popular optimization technique and many effective solvers have been developed. In this paper we use CPLEX solver, which is part of AIMMS[2] system, to estimate the optimal solution from the Eq.(4).

4 Experiments

In this section, we perform evaluations of the proposed method. The data sets we used in the experiments are described in the first part. After that, experimental results are given and detailedly described in the following sections.

4.1 Dataset and Evaluation Metric

There are almost no publicly available datasets with manually annotated gold standard keyphrases for news, due to the high expense of labor and time for manual annotation. In this experiment, we randomly selected 150 groups of online news articles from Goolge News. Three annotators participated in the annotation task. They were asked to manually assign keyphrases for each group of news. The keyphrases which at least two annotators have agreed on are selected as the "Golden" ones. Statistics on the dataset are shown in Table 1. The corpus data is divided into development set and test set. The development set, which contains 50 groups of news, is used to tune the parameters. The other 100 groups of news are used as test set.

We regard an extracted keyphrase as "correct" if it matches one of the ground truth. We measure the performance by Precision (the percentage of correct extracted keyphrases out of all the extracted ones), Recall (the percentage of correct extracted keyphrases out of the ground truth) and F-Measure (the harmonic mean of the precision and recall).

Description	value
# News articles	1103
# Words	345K
# News articles per group	7.35
# Labeled keyphrases per group	5.83

Table 1: Statistics on the dataset

4.2 Comparisons with Other Methods

Since the dataset used in this paper is manually labeled by ourselves, we implement three baseline methods on the same dataset for comparison.

BL-1: The titles of news articles provide a reasonable summary or keyphrase sequence. So baseline 1 is performed based on the titles of news articles. We sort the phrases in multi-news titles according to the TF·IDF scores and select top-k as keyphrases. We assign K to 6 after tuning the parameter.

BL-2: Many existing methods converted the keyphrase extraction as a classification problem. In this paper, we used SVM[3] as baseline 2. The features include TF·IDF, "First occurrence", and "Is in title or not". Those feature sets are similar to our objective function. We divided the dataset into five subsets and conducted a 5-fold cross-validation.

BL-3: We re-implemented the ranking approach proposed by Jiang et al. (2009) as baseline 3. This method employed Ranking SVM (Joachims, 2006), the learning to rank method, to perform keyphrase extraction. Feature sets are the same as the feature sets used in the BL-2. We also conducted a 5-fold cross-validation.

We used the following default values for the parameters of our method: $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.3$, $\mu = 0.1$, $\nu = 0.05$, $\zeta = 0.005$, $\xi = 16.5$, and $K = 6$. The meaning of these parameters are described in the previous section. And how to learning the optimal values will be discussed in section 4.4. The test set is used in this experiment. Since the average number of manual-

[2] http://www.aimms.com/

[3] SVM^{light} is used in our experiments, which can be downloaded from http://www.cs.cornell.edu/People/tj/svm_light.

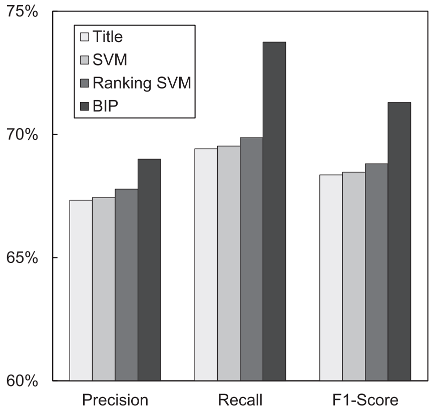

Figure 1: Comparison results of Title, SVM, Ranking SVM and our BIP-based methods .

labeled keyphrases is six, we selected the top 6 ones as keyphrases in all three baseline methods.

Figure 1 shows the performance comparison of BIP-based method with baselines. From the figure, we have the following observations. Firstly, BIP-based method consistantly outperforms all baselines under all evaluation metrics – Precision, Recall, and F1-Score. This indicates the robustness and effectiveness of our method. Furthermore, compared with the supervised methods, BIP-based method does not need any labeled corpus. Secondly, Ranking SVM performs slightly better than SVM. This is congruence with the previous conclusion given by Jiang et al. (2009). However, the improvement of BL-3 over BL-2 is not significant. We also observe that the performances of BL-1 are quite good. The precision, recall, and F1-score achieved by it are comparable with results of SVM and Ranking SVM.

4.3 Contribution of Constraints and Objective Function

To determine the contribution of different components of objective function and individual constraints, we omit components and constraints one by one to identify its contribution to the performance. Table 2 shows the results on development set. The first row represents the performance of the BIP-based method with all constraints and objective function with all three components. The parameters are default ones listed in the previous section.

The contribution of different components in the objective function is shown from the second row to fourth row. From the results we can observe that, TF·IDF is the most important feature in the objective function. Without the feature of TF·IDF, the evaluation metrics drop sharply from 72.68% to 59.82%. The candidate occurs in the title is also an important feature. It is consistent with the observations given by the results of BL-1. It gives about 17.27% relative improvement over the performance without it. Compared with the two features, the occurrence in the first sentence gives less contribution. The improvement given by it is not significant.

Table 2: Contribution of different components of objective function (TFIDF, InTitle, InFirstSenetence) and two constraints (Coverage and Coherence) under Precision, Recall, and F1-Score.

	Pre.	**Rec.**	**F.**
All	**71.45%**	**73.96%**	**72.68%**
All - TF·IDF	58.86%	60.82%	59.82%
All - InTitle	60.96%	62.77%	61.85%
All - InFirstSentence	71.00%	73.20%	72.08%
All - Coverage	68.56%	70.68%	69.60%
All - Coherence	70.67%	72.85%	71.74%

The fifth row shows the results without the coverage constraint. From the result, we observe that the coverage constraint is effective, which can give more than 4.2% relative improvement. The contribution of coherence constraint is shown in the end of the table. Althoug its contribution is less than that of coverage constraint, an outlier keyphrase may highly impact the user experience. Coherence constraint is important to improve user experience.

4.4 Varying Parameters

As we mentioned in the previous section, there are eight parameters which should be adjusted in our method. One may concern the problem of parameters tuning. In order to answer this question, in this section, we explore the impact of different parameters on our approach's performance in the development set. Except the parameter under investigation, the other parameters are set to the default values which are listed in the Section 4.2.

4.4.1 Number of Keyphrases K

Figure 2 presents the performance varying the number of keyphrases, which ranges from 1 to 10. K is one of the most important arguments leading to the trade-off between precision and recall. Larger K increases recall but decreases precision. From this figure, we observe that the best

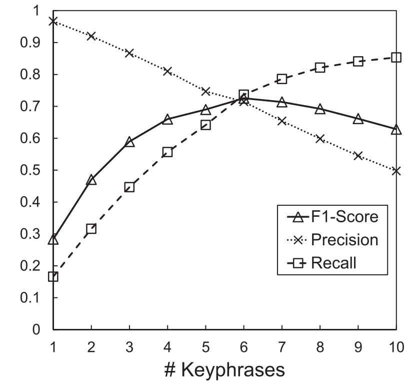

Figure 2: Results of varying the number of extracted keyphrases using the proposed BIP-based extraction method.

result is achieved at the point $K = 6$, which is similar to the average number of manually selected keyphrases. We also observe that the F1-Score drops quickly when K is bigger than 7. The main reason is that only a small number of phrases which should be selected are ranked after the top 10.

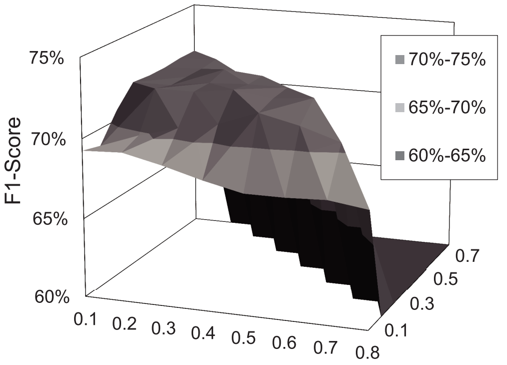

Figure 3: Results of varying the parameters α, β, γ in the objective function.

4.4.2 α, β, γ in the Objective Function

In the objective function, there are three parameters α,β, and γ , which are used to trade off among TF·IDF and two locality features. Figure 3 gives the F1-Score surface varying α and β. Since $\alpha + \beta + \gamma$ equals to 1, α and β are used as x-axis and y-axis in the figure. We have found that the surfaces are almost concave around a number of areas. Therefore, a simple hill-climbing search can be used to optimize F1-Score. Since the surface is almost concave, the global maximum can be easily achieved though a few initial seeds. For example, the optimal parameters for this experiment are $\alpha = 0.4$, $\beta = 0.3$. The γ can be calculated through function $1 - \alpha - \beta$.

Table 3: Influence of the Coverage threshold ζ

ζ	**Pre.**	**Rec.**	**F.**
0.001	69.44%	71.59%	70.50%
0.002	70.33%	72.50%	71.40%
0.003	71.22%	73.43%	72.31%
0.004	71.00%	73.20%	72.08%
0.005	**71.45**%	**73.65**%	**72.53**%
0.006	71.33%	73.54%	72.42%
0.007	71.22%	73.43%	72.31%
0.008	71.11%	73.31%	72.19%
0.009	70.67%	72.85%	71.74%

4.4.3 Coverage threshold ζ

Coverage threshold ζ represents the property that the extracted keyphrases should cover most of the important aspects of a news event. We want the aspect coverage for all topics to exceed the threshold. Table 3 shows the results when ζ ranges from 0.001 to 0.009. All of them perform better than the result without coverage constraint, and the best result is achieved at $\zeta = 0.005$. From the results, we observe that the coverage threshold ζ can also be easily selected. From 0.005 to 0.008, the changes of F1-score are not significant. When the coverage threshold is above 0.02, in order to get the solution of the ILP program, the impact of objective function would be limited. We think that it is the main reason of why best result is achieved at a small value threshold.

4.4.4 Coherence threshold ξ

Finally, we explore the inference of ξ, which is used to represent the word coherence. When the threshold is below 12, there would be too many coherence constraints. More than 30.05% word pairs can satisfy the threshold. Under this condition, no solution can be estimated in some cases. When the threshold is above 32, there are rarely word pairs satisfying the threshold. In other words, there would be no coherence constraints. In table 4 we show the influence of ξ, which ranges from 15 to 18. Similar to the results of coverage threshold, a large range of ξ's value can achieve satisfactory result. $\xi = 16.5$ achieves the best result 72.53%.

Table 4: Influence of the Coherence threshold ξ

ξ	Pre.	Rec.	F.
15.0	68.00%	70.10%	69.04%
15.5	69.89%	72.05%	70.95%
16.0	70.78%	72.97%	71.85%
16.5	**71.45%**	**73.65%**	**72.53%**
17.0	71.22%	73.43%	72.30%
17.5	71.00%	73.20%	72.08%
18.0	71.00%	73.20%	72.08%

4.5 Extracting Example

Table 5 shows examples of extracted keyphrases by different methods from a group of news articles about "Master Kong applies for TDR listing in Taiwan". Top 6 extracted keyphrases for each method are shown in the table, and the correct ones are marked with "(+)". From table 5, we observe that keyphrases extracted through BIP-based method are relevant, coherent, with good coverage. Without the coverage constraint, "Taiwan Depositary Receipt" and it's abbreviation "TDR" are both selected. And, the topic coverage cannot be well satisfied through the top keyphrases. For SVM and Ranking SVM, they separately consider each word, some of the high frequency words are selected as keyphrases, such as "billion", and "issue". However, those words are not meaningful.

5 Conclusions

In this paper, we have presented a novel keyphrase extraction approach. It adapts the integer linear programming methods to the keyphrase extraction problem by casting features and criteria as objective function and constraints.

By integrating TF·IDF and two locality features as objective function, and the coverage and coherence properties as constraints, the proposed ILP-based unsupervised approach achieves better performance than the state-of-the-art supervised approaches, SVM and Ranking SVM. Contributions of constraints and different components of the object function are experimental evaluated. In the objective function, the TF·IDF is the most important feature. Locality features can further improve the performance. Results also demonstrate that both the coverage and coherence constraints are useful to keyphrase extraction task. We also detail the impact of parameters used in our approach. Through experimental results, we demonstrate that the parameters are not sensitive. The value of them can be easily estimated using simple hill-climbing search methods.

Table 5: Example of extracted keyphrases by SVM, Ranking SVM and BIP-based method*.

BIP
Masker Kong(+), Ting Hsin International Group(+), Taiwan Depositary Receipt(+), instant noodle, Taiwan Stock Exchange(+), NT$30 billion
BIP without Coverage constraint
Masker Kong(+), Taiwan Depositary Receipt(+), TDR, lunch, Taiwan Stock Exchange(+), Taiwan
BIP without Coherence constraint
Masker Kong(+), Taiwan Depositary Receipt(+), Taiwanese-invested food producer(+), IPO, issue, Taiwan Stock Exchange(+)
SVM
Taiwan Depositary Receipt(+), Masker Kong(+), China market, TDR, Hong Kong Exchanges, billion
Ranking SVM
Masker Kong(+), China market, TDR, Taiwan Depositary Receipt(+), issue, Taiwan

*The keyphrases are translated from Chinese.

Acknowledgments

The author wishes to thank the anonymous reviewers for their helpful comments. This work was partially funded by 973 Program (2010CB327906), National Natural Science Foundation of China (61003092, 61073069), Shanghai Science and Technology Development Funds(10dz1500104), Doctoral Fund of Ministry of Education of China (200802460066), Shanghai Leading Academic Discipline Project (B114), and Key Projects in the National Science & Technology Pillar Program(2009BAH40B04).

References

Dimitris Alevras and Manfred W. Padberg. 2001. *Linear Optimization and Extensions: Problems and Solutions*. Springer.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March.

Mo Chen, Jian-Tao Sun, Hua-Jun Zeng, and Kwok-Yan Lam. 2005. A practical system of keyphrase extraction for web pages. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, pages 277–278, New York, NY, USA. ACM.

Eibe Frank, Gordon W. Paynter, Ian H. Witten, Carl Gutwin, and Craig G. Nevill-Manning. 1999. Domain-specific keyphrase extraction. In *Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2*, pages 668–673, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Anette Hulth. 2003. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing - Volume 10*, pages 216–223, Morristown, NJ, USA. Association for Computational Linguistics.

Xin Jiang, Yunhua Hu, and Hang Li. 2009. A ranking approach to keyphrase extraction. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 756–757, New York, NY, USA. ACM.

Thorsten Joachims. 2006. Training linear svms in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 217–226, New York, NY, USA. ACM.

Zhiyuan Liu, Peng Li, Yabin Zheng, and Maosong Sun. 2009. Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 257–266, Morristown, NJ, USA. Association for Computational Linguistics.

André F. T. Martins, Noah A. Smith, and Eric P. Xing. 2009. Concise integer linear programming formulations for dependency parsing. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1*, ACL-IJCNLP '09, pages 342–350, Morristown, NJ, USA. Association for Computational Linguistics.

Olena Medelyan and Ian H. Witten. 2006. Thesaurus based automatic keyphrase indexing. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '06, pages 296–297, New York, NY, USA. ACM.

Olena Medelyan, Eibe Frank, and Ian H. Witten. 2009. Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1318–1327, Morristown, NJ, USA. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 404–411, Barcelona, Spain, July. Association for Computational Linguistics.

Dan Roth and Wen-tau Yih. 2005. Integer linear programming inference for conditional random fields. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pages 736–743, New York, NY, USA. ACM.

T.D.Nguyen and M.-Y. Kan. 2007. Keyphrase extraction in scientific publications. In *Proceedings of International Conference on Asian Digital Libraries*.

Takashi Tomokiyo and Matthew Hurst. 2003. A language model approach to keyphrase extraction. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*, pages 33–40, Morristown, NJ, USA. Association for Computational Linguistics.

Peter D. Turney. 2000. Learning algorithms for keyphrase extraction. volume 2, pages 303–336, Hingham, MA, USA, May. Kluwer Academic Publishers.

Peter D. Turney. 2003. Coherent keyphrase extraction via web mining. In *Proceedings of the 18th international joint conference on Artificial intelligence*, pages 434–439, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Xiaojun Wan and Jianguo Xiao. 2008. Single document keyphrase extraction using neighborhood knowledge. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 2*, pages 855–860. AAAI Press.

Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig G. Nevill-Manning. 1999. Kea: practical automatic keyphrase extraction. In *Proceedings of the fourth ACM conference on Digital libraries*, DL '99, pages 254–255, New York, NY, USA. ACM.

Kristian Woodsend and Mirella Lapata. 2010. Automatic generation of story highlights. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 565–574, Morristown, NJ, USA. Association for Computational Linguistics.

Y.Matsuo and M.Ishizuka. 2004. Keyword extraction from a single document using word co-occurrence statistical information. In *International Journal on Artificial Intelligence Tools*.

Hongyuan Zha. 2002. Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '02, pages 113–120, New York, NY, USA. ACM.