Effectively Leveraging Entropy and Relevance for Summarization

Wenjuan Luo[1,2], Fuzhen Zhuang[1,2], Qing He[1], and Zhongzhi Shi[1]

[1] The Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China

[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

{luowj,zhuangfz,heq,shizz}@ics.ict.ac.cn

Abstract. Document summarization has attracted a lot of research interest since the 1960s. However, it still remains a challenging task on how to extract effective feature for automatic summarization. In this paper, we extract two features called entropy and relevance to leverage information from different perspectives for summarization. Experiments on unsupervised and supervised methods testify the effectiveness of leveraging the two features.

Keywords: summarization, entropy, relevance, sentence feature extraction.

1 Introduction

With the actual huge and continuously growing of World Wide Web, the amount of information in the public domain grows explosively. As a result, there is a vast demand for new technologies that can effectively process information. Document summarization is an essential technology to overcome this obstacle in technological environments [1].

The main motivation of document summarization is to help users capture the major topics of a document with less effort [2]. Different summarization tasks make the query-oriented summary different from the generic summary [3]. Besides, document summarization can be categorized as abstract-based and extract-based summaries [1]. In this paper, we aim to generate extract-based generic summary, i.e., to select a combination of sentences which are the most important for the overall understanding of the document[4].

Many algorithms, supervised and unsupervised, have been applied for document summarization. Jones [5] gives a review of the research on automatic summarization over the last decade. Based on our study, we argue that a good summary should be compact while cover as many aspects in the document as possible. We propose a feature called entropy to measure the coverage from the inner-sentence level, and another feature called relevance to indicate the compactness from the intra-sentence level:

Entropy–This feature denotes the quantity of information implied by the sentence. As you may notice, long sentences are likely to cover more aspects

P.-J. Cheng et al. (Eds.): AIRS 2010, LNCS 6458, pp. 241–250, 2010.

in the document than short sentences. Note that a long sentence usually has a comparably larger entropy than a short sentence. Hence, a large entropy of sentence possibly implies a large converge.

Relevance–This feature measures the intra-sentence relationships between sentences. On the whole, sentences sharing a considerable number of words with other sentences often have high relevances. Therefore, choosing sentences correlated with each other probably leads to a compact summary.

As you may notice, leveraging the above two features may generate summaries that prefer long sentences which are strongly related to other sentences in the document. We aim to balance the importance within the sentence and between the sentences to generate compact summaries covering as many aspects of the document as possible.

In order to testify the effectiveness of leveraging the two features, or say heuristic rules[6], we firstly score sentences on the strength of these two features and perform a simple sentence selection method for unsupervised summarization. Furthermore, we combine the two features with other features [2,7] extracted from sentences and apply regression methods for supervised summarization.

We evaluate the performance of our methods from DUC01(http://duc.nist.gov) on an open benchmark data set. Experiments show that leveraging the two proposed heuristic rules are contributive to summary generation from both the unsupervised and the supervised perspective.

The rest of the paper is organized as follows: Section 2 introduces the preliminary knowledge and gives the problem definition, Section 3 presents the details of the algorithms, Section 4 gives experimental results and comparisons of our methods with baseline methods, Section 5 describes the related work, and we conclude our paper in Section 6.

2 Preliminary Knowledge

2.1 Feature Extraction

Since sentences could be viewed as a vector of words, in order to analyze the significance within the sentence quantitatively, we adopt entropy as a metric:

$$I(x_i) = -\sum_{j \in x_i} p_{ij} \cdot \log(p_{ij}) \tag{1}$$

where x_i denotes the i-th sentence, $I(x_i)$ denotes the amount of information covered by the i-th sentence, p_{ij} denotes the probability of the j-th word in the i-th sentence. From the above formula, we can see that a longer sentence is more likely to have a larger entropy. Therefore, a sentence of large entropy is very likely to cover more aspects of the document than a sentence of small entropy.

Moreover, we extract another feature called relevance to show the compactness by calculating the relationship of a sentence between other sentences:

$$R(x_i) = \sum_{j \neq i} S(x_j, x_i),\ S(x_j, x_i) = Overlap(x_j, x_i)/length(x_i) \tag{2}$$

where $R(x_i)$ denotes the relevance of the i-th sentence, $S(x_j, x_i)$ is the similarity between the j-th sentence and the i-th sentence based on directed backward graph according to Mihalcea's experiments[8], $Overlap(x_j, x_i)$ denotes the number of words co-occurring in the j-th and the i-th sentence, and $length(x_i)$ is the length of the i-th sentence. Hence, sentences with high relevance probably compose more compact summaries than sentences without.

Besides, for regression methods, we extract entropy and relevance together with other typical features from the document as [2,7] in Table 1.

Table 1. Features for Supervised Summarization

f1	the position of the sentence	f2	the length of the sentence
f3	the likelihood of the sentence	f4	the number of thematic words
f5	the number of low frequency words	f6	the LSA-Score of the sentence
f7	the number of 2-gram keywords	f8	number of words appearing in other sentences
f9	*the entropy of the sentence*	f10	*the relevance of the sentence*

2.2 Problem Formulation

Given a document $x = \{x_1, x_2, ..., x_N\} \in \mathcal{X}$, where x is a document, x_i represents the i-th sentence, and $\mathcal{X}$ denotes the space of all the documents.

For a single document, our target is to extract the most representative $y = \{s_1, s_2, ..., s_k\}$ as the final summary, where s_i represents the i-th sentence in the summary.

Unsupervised Summarization. As discussed in Section 2.1, a large $I(s_i)$ of sentence probably suggests a large coverage, and large $R(s_i)$s indicate that sentences are relevant which probably leads to compact summaries. Therefore, in order to generate compact summaries with a large coverage, we aim to maximize the following objective function:

$$\mathbb{S}^* = argmax \sum_{s_i \in \mathbb{S}} \{\alpha \cdot I(s_i) + \beta \cdot R(s_i)\} \tag{3}$$

where $\mathbb{S}^*$ denotes the summary. Take a deep look at Equation (3), substitute the $I(s_i)$ and $R(s_i)$ with Equation (1) and (2), we have:

$$\begin{aligned} F(\mathbb{S}) &= \sum_{s_i \in \mathbb{S}} \{\alpha \cdot I(s_i) + \beta \cdot R(s_i)\} \\ &= \sum_{s_i \in \mathbb{S}} \{-\alpha \cdot \sum_{j \in s_i} p_{ij} \cdot \log(p_{ij}) + \beta \cdot \frac{\sum_{x_m \neq s_i} Overlap(s_i, x_m)}{length(s_i)}\} \\ &= \sum_{s_i \in \mathbb{S}} \{-\alpha \cdot \sum_{j \in s_i} p_{ij} \cdot \log(p_{ij}) + \beta \cdot \log(e^{\frac{\sum_{x_m \neq s_i} Overlap(s_i, x_m)}{length(s_i)}})\} \\ &= \sum_{s_i \in \mathbb{S}} \{\log(\prod_{j \in s_i} p_{ij}^{-\alpha p_{ij}} \cdot \prod_{x_m \neq s_i} e^{\frac{\beta Overlap(s_i, x_m)}{length(s_i)}})\} \end{aligned} \tag{4}$$

where x_m is the m-th sentence in the document. After removing stopwords, the number of words in sentences is small, while the effect of $R(s_i)$ grows exponentially as the number of sentences increases. Therefore, if we use relevance in its original form, the $R(s_i)$ will be dominant in the objective function. For a better balance, we adjust our objective function as follows:

$$F(\mathbb{S}) = \sum_{s_i \in \mathbb{S}} \{\alpha \cdot I(s_i) + \beta \cdot \log(R(s_i))\} \tag{5}$$

Supervised Summarization. Given a document $x = \{x_1, x_2, ..., x_N\} \in \mathcal{X}$, for each sentence x_i, there is a corresponding y_i indicating whether the i-th sentence is in the summary. Now assume we have all the training data preprocessed as:

$$\{x_i^{(1)}, x_i^{(2)}, ..., x_i^{(j)}, ..., x_i^{(n)}, y_i | i = 1, 2, ..., n\} \tag{6}$$

where $x_i^{(j)}$ denotes the j-th feature of the i-th sentence,$\{x_i^{(1)}, x_i^{(2)}, ..., x_i^{(j)}, ..., x_i^{(n)}\}$ is the feature vector of x_i, for a classification problem, y_i is either 0 or 1, and for regression problem, y_i denotes the ROUGE-2-P score of the i-th sentence, which will be described in Section 4.2. We aim to construct a discriminant function such that:

$$\mathcal{F}(\mathcal{X}, \mathcal{Y}) : \Psi^T \cdot \mathcal{X} = \mathcal{Y} \tag{7}$$

We attempt to find the Ψ^T with the least generalization error through the training process. The predicted summary y would be extracted according to the following equation:

$$y = argmax\{\Psi^T \cdot x\} \tag{8}$$

The sentences with the highest values of y will constitute the final summary.

3 Proposed Methods

3.1 Entropy-and-Relevance-Based Summarization

As we have given the objective function in Section 2.2, in order to select the most important sentence in the sense of entropy and relevance. As we can see, to maximize the objective function is to find the sentences with highest $I(s_i) + (\beta/\alpha) \cdot \log(R(s_i))$.

Algorithm 1. Entropy and Relevance based Summarization

Input: The term frequency matrix of each document, the number of sentences (k) to be selected as summary, the weight of entropy(α), the weight of relevance(β).
Output: The sequence of sentences selected as final summary.

Step 1: Given the term frequency matrix, construct the backward similarity matrix.
Step 2: For each sentence s_i, calculate its entropy($I(s_i)$) and relevance($R(s_i)$).
Step 3: Calculate the score of $I(s_i) + (\beta/\alpha) \cdot \log(R(s_i))$ for each sentence.
Step 4: Output the top k sentences with highest scores as the final summary.

For a better balance between entropy and relevance, we adopt different α and β to see the effect of leveraging the two features. Specifically, for simplicity, we consider the parameter tuning to be 1 and β/α. Experimental results will be given in Section 4.

3.2 Regression-Based Summarization

From the supervised perspective, we adopt the linear regression [9] and Extreme Learning Machine (ELM) [10] regression to leverage the entropy and relevance. For linear regression, we simply apply the following model:

Given the whole document space $X = \{x_1, x_2, ..., x_N\}$, the feature vector $x_i = (x_i^{(1)}, x_i^{(2)}, ..., x_i^{(n)})$ and its corresponding y_i, for linear regression, we have

$$f(x_i) = \beta_0 + \sum_{j=1}^{n} x_i^{(j)} \beta_j \tag{9}$$

We use the minimum squared errors to estimate β, therefore, we have:

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{n} x_i^{(j)} \beta_j)^2 \tag{10}$$

In Equation (10), to minimize the RSS with respect to β,

$$\frac{\partial RSS}{\partial \beta} = -2X^T(y - X\beta), \frac{\partial^2 RSS}{\partial \beta \partial \beta^T} = -2X^T X \tag{11}$$

Assume X is full column rank, let $\frac{\partial RSS}{\partial \beta} = 0$, we can obtain a solution of β

$$\beta^* = (X^T X)^{-1} X^T y \tag{12}$$

On the other hand, the ELM regression differs from linear regression in the way how β is obtained. For the single hidden layer feedforward networks (SLFNs) with M hidden neurons, the problem could be modeled as follows:

$$\sum_{i=1}^{N} \|o_i - y_i\| = 0 \text{ , where } o_i = \sum_{j=1}^{M} \beta_j g(w_j x_i^{(j)} + b_j), i = 1, ..., N \tag{13}$$

In Equation (13), the goal of the $g(\cdot)$ function is to approximate all the training data with zero means. Therefore, for all the N training sample, the goal is to minimize the cost function:

$$E = \sum_{i=1}^{N} (\sum_{j=1}^{M} \beta_j g(w_j x_i^{(j)} + b_j) - y_i)^2 \tag{14}$$

Under the ELM model, Huang et al. [10] formularized the problem as:

$$H\beta = Y \text{ where}$$

$$H(w_1, ..., w_M, b_1, ..., b_M, x_1, ..., x_N) = \begin{pmatrix} g(w_1 x_1^{(1)} + b_1), \cdots, g(w_M x_1^{(M)} + b_M) \\ \vdots \qquad \cdots, \qquad \vdots \\ g(w_1 x_N^{(1)} + b_1), \cdots, g(w_M x_N^{(M)} + b_M) \end{pmatrix}$$

$$\text{and } \beta = [\beta_1^T, \cdots, \beta_M^T]^T, Y = [y_1^T, \cdots, y_N^T]^T \tag{15}$$

According to Equation (15), through the use of the Moore−Penrose generalized inverse, a solution of β is [11]:

$$\beta^* = H^\dagger Y \tag{16}$$

where, $H^\dagger$ is the Moore−Penrose generalized inverse of H. Detail steps of deduction could be found in [10] and [11].

For both kinds of regression, the predictive summary of x could be calculated through:

$$y^* = x\beta^* \tag{17}$$

For each document, we select the sentences with the highest values of y^* into the final summary. As we have mentioned, the goal of using regression methods is to learn how to balance different features through the training procedure. Effectiveness of the leveraging is shown in the next section.

4 Experiments

4.1 Experimental Setting

Experimental Data: The DUC(Document Understanding Conference) 2001 data set is used to evaluate the effectiveness of leveraging the two features. The whole data set includes 147 documents of 6921 sentences together with the corresponding ground-truth summaries.

Baseline Methods: For unsupervised methods, we implement the Luhn's method, RANDOM[7], LSA and HITS to compare with our entropy and relevance based summarization. For supervised methods, the Linear Regression and ELM regression are applied based on different feature sets for evaluation.

Evaluation Metric: The ROUGE evaluation toolkit [12] adopted by DUC for automatic summarization evaluation is highly correlated with human evaluations. In this paper, we employ this toolkit to evaluate the performance of our proposed methods. There are several kinds of ROUGE metrics, here we introduce the most commonly used sub-metrics as follows:

1. ROUGE-N-R is the recall rate of summary from the n−gram point of view. It can be calculated as follows [12]:

$$\text{ROUGE-N-R} = \frac{\sum_{s\in y^*}\sum_{gram_n\in s} Count_{match}(gram_n)}{\sum_{s\in y}\sum_{gram_n\in s} Count(gram_n)}$$

2. ROUGE-N-P is the precision rate summary from the n−gram point of view. It can be calculated as follows :

$$\text{ROUGE-N-P} = \frac{\sum_{s\in y^*}\sum_{gram_n\in s} Count_{match}(gram_n)}{\sum_{s\in y^*}\sum_{gram_n\in s} Count(gram_n)}$$

3. ROUGE-N-F is the F_1 metric of ROUGE-N-R and ROUGE-N-P and could be calculated as follows:

$$\text{ROUGE-N-F} = \frac{2 * \text{ROUGE-N-R} * \text{ROUGE-N-P}}{\text{ROUGE-N-R} + \text{ROUGE-N-P}}$$

In the above equations, N is the length of words in n-gram, s is the sentence in summary, y^* denotes the generated summary and y is the ground−truth summary. $Count_{match}(gram_n)$ is the number of n-gram co-occurring between y^* and y. $Count(gram_n)$ denotes the occurrence number of n-gram words in the corresponding summary.

4.2 Performance Evaluation

Results of Unsupervised Methods: We implement several unsupervised methods for summarization to compare with our ERBS summarization. The results are shown in Table 2.

Table 2. Results of Unsupervised Methods

Name	ROUGE-1-R	ROUGE-1-P	ROUGE-1-F	ROUGE-2-R	ROUGE-2-P	ROUGE-2-F
luhn	0.514	0.510	0.512	0.355	0.353	0.354
RANDOM	0.524	0.552	0.535	0.408	0.428	0.416
lsa	0.506	0.501	0.503	0.351	0.349	0.351
HITS	0.575	0.576	0.575	0.451	0.453	0.452
ERBS	**0.601**	**0.597**	**0.599**	**0.481**	**0.479**	**0.480**

From Table 2, we can find that ERBS achieves the best performance among the unsupervised methods and gains a significant improvement over all baseline methods. This shows the effectiveness of the proposed heuristic rules– entropy and relevance. Moreover, we give the detail results of ERBS with different parameters in Fig. 1.

From the left figure in Fig. 1, we can see that ROUGE-scores changes with different gama, where gama=β/α. From the figure on the right, we can see the best result is obtained when gama=1.2, as is displayed in Table 2. When gama=0, the ERBS just selects sentences with the highest entropy, as gama grows, the ERBS pays more and more attention to relevance. The result of ERBS converges to selecting sentences with the highest relevance when gama grows infinite. As we can see from the figures, leveraging entropy and relevance achieves better performance than emphasizing one single heuristic rule.

Results of Supervised Methods. To further investigate the effect of entropy and relevance, we also listed the compared results of LR, LR+E, LR+R, LR+E+R, ELM, ELM+E, ELM+R and ELM+E+R in Table 3. LR denotes Linear-Regression dealing with features excluding entropy and relevance, while LR+E includes entropy, LR+R includes relevance and LR+E+R includes all the features. The methods based on ELM-regression are named similarly.

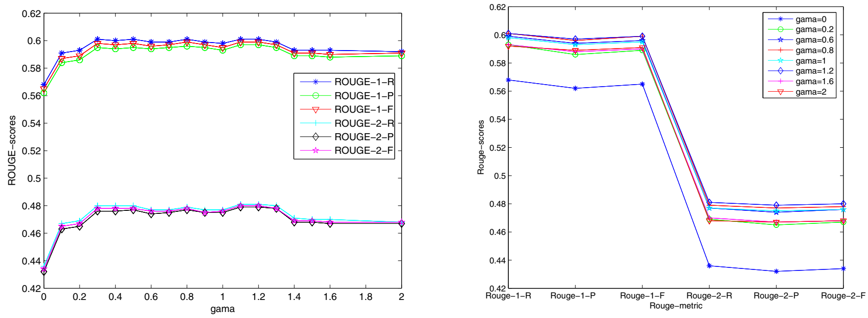

Fig. 1. Rouge-scores of ERBS according to different gama

For a better comparison, we also employ the weighted longest common subsequence (ROUGE-W). And the weight parameter is set to be 1.2 in the experiments. Comparing the LR+E+R with LR, LR+E and LR+R, we can see that LR+E+R performs the best. And it is the same with the ELM. This suggests that leveraging the two features enhances the regression based summarization.

Table 3. Results of Regression Methods

ROUGE-Score	LR	LR+E	LR+R	LR+E+R	ELM	ELM+E	ELM+R	ELM+E+R
ROUGE-1-R	0.54421	0.61309	0.60750	0.61750	0.55229	0.61121	0.60822	**0.62139**
ROUGE-1-P	0.55679	0.60905	0.60333	0.61326	0.55483	0.60670	0.60351	**0.61679**
ROUGE-1-F	0.54926	0.61077	0.60512	0.61510	0.55324	0.60866	0.60556	**0.61879**
ROUGE-2-R	0.42278	0.49969	0.49140	0.50651	0.42557	0.49536	0.49614	**0.51123**
ROUGE-2-P	0.43095	0.49701	0.48867	0.50344	0.42760	0.49234	0.49288	**0.50813**
ROUGE-2-F	0.42615	0.49810	0.48980	0.50473	0.42635	0.49360	0.49427	**0.50943**
ROUGE-W-R	0.22252	0.25208	0.24867	0.25433	0.22445	0.25121	0.25077	**0.25749**
ROUGE-W-P	0.43509	0.48042	0.47382	0.48428	0.43140	0.47752	0.47654	**0.48973**
ROUGE-W-F	0.29333	0.32989	0.32540	0.33274	0.29453	0.32846	0.32786	**0.33673**

5 Related Work

Luhn [13] firstly proposed automatic summarization and addressed the significance of sentences for sentence extraction in the 1960s. Moreover, LSA (Latent Semantic Analysis) was applied to identify semantically important sentences for summarization based on the Singular Vector Decomposition(SVD) of matrix[14] one decade ago. Recently, Lee et al. [1] proposed an unsupervised document summarization method using the Non-negative Matrix Factorization (NMF).

Also, ranking algorithms such as HITS and PageRank could be applied for summarization since documents could be represented as graphs [6]. Mihalcea [8] implemented HITS based on undirected graph, forward directed graph and backward directed graph. Among all the graphs, the method based on the backward directed graph performs best.

Meanwhile, supervised algorithms have also been applied for document summarization. Generally, these methods first extract a set of features from the document, and then train a summarizer to predict whether a sentence should be selected into the summary [6]. Such features include linguistic features and statistical features, such as rhetorical structure [15], the position of the sentence in the document, the sentence length, and so on [2,16]. Some complex features are included in order to improve the performance of supervised algorithms such as, the LSA score of the sentence, the HITS score of the sentence [3], the PageRank score of the sentence [7], and so on.

Learning methods such as Naive Bayes (NB), C4.5 [16], Logistic Regression (LR), and Neural Network (NN) could be applied to train summarizers based on features extracted from sentences [2]. Rijsbergen et al. [17] introduces an SVM-based method aiming at constructing a decision boundary between summary and non-summary sentences. Moreover, HMM(Hidden Markov Models) was proposed by Conroy and O'Leary [3] in 2001 based on three extracted features.

However, HMM could not fully exploit linguistic features of sentences since its independence assumption. Shen et al. [2] proposed Conditional Random Fields(CRF) for summarization, which considers the summarization task as a sequence labeling problem. In order to enhance diversity, coverage and balance for summarization, Li et al. [7] proposed another supervised method through a structure learning framework.

For supervised methods, whether the extracted features indicate the useful information for summarization strongly affects the quality of summaries. Moreover, the capability of the adopted method also influences the quality of summarization.

6 Conclusion

To generate compact summaries with possibly large coverage, we extract entropy and relevance from sentences for summarization. Then we perform unsupervised summarization named ERBS and supervised summarization utilizing Linear Regression and ELM regression in order to validate the effectiveness of leveraging the two features. Experimental results show that the ERBS outperforms other baseline unsupervised methods. Moreover, the results of linear regression and ELM regression based summarization also indicate that leveraging the two features is beneficial for automatic summarization.

Acknowledgements

This work is supported by the National Science Foundation of China (No.60675010, 60933004, 60975039), 863 National High-Tech Program (No.2007AA01Z132), National Basic Research Priorities Programme (No.2007CB311004) and National Science and Technology Support Plan (No.2006BAC08B06).

References

1. Lee, J.-H., Park, S., Ahn, C.-M., Kim, D.: Automatic generic document summarization based on non-negative matrix factorization. Information Processing and Management 45, 20–34 (2009)
2. Shen, D., Sun, J.-T., Li, H., Yang, Q., Chen, Z.: Document summarization using conditional random fields. In: International Joint Conference On Artificial Intelligence, pp. 2862–2867 (2007)
3. Conroy, J.M., O'Leary, D.P.: Text summarization via hidden markov models. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 406–407. ACM Press, New York (2001)
4. Aliguliyev, R.M.: A new sentence similarity measure and sentence based extractive technique for automatic text summarization. IExpert Systems with Applications 36, 7764–7772 (2009)
5. Jones, K.S.: Automatic summarizing: The state of the art. Information processing and Management 43, 1449–1481 (2007)
6. Shen, D., Yang, Q., Chen, Z.: Noise reduction through summariation for web-page classification. Information Processing and Management 42, 1735–1747 (2007)
7. Li, L., Zhou, K., Xue, G.-R., Zha, H., Yu, Y.: Enhancing diversity, coverage and balance for summarization through structure learning. In: WWW 2009: Proceedings of the 18th international conference on World wide web, vol. 4825, pp. 71–80. ACM Press, New York (2009)
8. Mihalcea, R.: Language independent extractive summarization. In: Proceedings of the 20th national conference on Artificial intelligence, vol. 4, pp. 1688–1689 (2005)
9. Montgomery, D.C., Peck, E.A., Vinin, G.G.: Introduction to linear regression analysis, 2nd edn. Wiley, Chichester (1992)
10. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: A new learning scheme of feedforward neural networks. In: International Joint Conference on Neural Networks 2004, vol. 2, pp. 985–990 (2004)
11. Matrices, D.S.: Theory and Applications, 1st edn. Springer, Heidelberg (2002)
12. Lin, C.Y., Hovy, E.: Automatic evaluation of summaries using n-gram cooccurrence statistics. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, pp. 71–78 (2003)
13. Luhn, H.P.: The automatic creation of literature abstracts. IBM Journal of Research and Development 2, 159–165 (1958)
14. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 19–25. ACM Press, New York (2001)
15. MarcuM, D.: From discourse structures to text summaries. In: ACL 1997 Workshop on Intelligent Scalable Text Summarization, pp.82–88 (1997)
16. Neto, J.L., Freitas, A.A., Kaestner, C.A.A.: Automatic text summarization using a machine learning approach. In: Bittencourt, G., Ramalho, G.L. (eds.) SBIA 2002. LNCS (LNAI), vol. 2507, pp. 205–215. Springer, Heidelberg (2002)
17. Rijsbergen, C.V.: Information Retrieval, 2nd edn. Butterworths (1979)