Information Processing and Management 45 (2009) 20–34

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Automatic generic document summarization based on non-negative matrix factorization

Ju-Hong Lee [a,*], Sun Park [b], Chan-Min Ahn [a], Daeho Kim [c]

[a] Department of Computer Science and Information Engineering, Inha University, Incheon, Republic of Korea
[b] Department of Computer Engineering, Honam University, Gwangju, Republic of Korea
[c] Department of Communication and Information, Inha University, Incheon, Republic of Korea

ARTICLE INFO

Article history:
Received 20 August 2007
Received in revised form 11 February 2008
Accepted 13 June 2008
Available online 8 August 2008

Keywords:
Generic summarization
NMF
LSA
Semantic feature
Semantic variable

ABSTRACT

In existing unsupervised methods, Latent Semantic Analysis (LSA) is used for sentence selection. However, the obtained results are less meaningful, because singular vectors are used as the bases for sentence selection from given documents, and singular vector components can have negative values. We propose a new unsupervised method using Non-negative Matrix Factorization (NMF) to select sentences for automatic generic document summarization. The proposed method uses non-negative constraints, which are more similar to the human cognition process. As a result, the method selects more meaningful sentences for generic document summarization than those selected using LSA.

1. Introduction

As the Internet sees increasingly wider use, the amount of information in the public domain continues to increase, rendering much of this information redundant. As such, new technologies that can process information efficiently are needed by users. Document summarization is an essential technology to overcome this obstacle in technological environments.

Automatic document summarization is the process of reducing the size of documents while presenting the important semantic content. Its purpose is to identify a summary of a document without reading the entire document. Content is extracted from an information source, and the most important content is presented to a user in a condensed form and in a manner appropriate for the user's or application's needs. This process can be used in many applications such as information retrieval, intelligence gathering, information extraction, text mining, and indexing (Luhn, 1958; Mani & Maybury, 1999; Mani, 2001).

Document summarization methods can be classified into generic summarization and query-based summarization. While generic summarization distills the summarized text and presents the important semantic content of given documents, query-based summarization presents the summaries that are closely related to the query (Marcu, 1999; Mani & Maybury, 1999; Mani, 2001).

Generally, a document is comprised of major and minor topics. A large number of the sentences in a document are related to the major topics. Some sentences, then, are related to minor topics, which supplement the major topics. A good generic summary should contain the major topics of the document and minimize redundancy (Gong & Liu, 2001).

* Corresponding author. Tel.: +82 32 860 7453; fax: +82 32 876 8052.
E-mail addresses: juhong@inha.ac.kr (J.-H. Lee), sunpark@honam.ac.kr (S. Park), ahnch1@datamining.inha.ac.kr (C.-M. Ahn), jinseon@inha.ac.kr (D. Kim).

0306-4573/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2008.06.002

Generally, automatic generic document summarization methods can be divided into two categories: supervised and unsupervised methods (Mani, 2001; Mani & Maybury, 1999). Supervised methods are based on algorithms that use a large amount of human-made summaries, and as a result, are most useful for documents that are relevant to the summarizer model. Thus, they do not necessarily produce a satisfactory summary for documents that are not similar to the model. In addition, when users change the purpose of summarization or the characteristics of documents, it becomes necessary to reconstruct the training data or retrain the model (Amini & Gallinari, 2002; Chuang & Yang, 2000; Nomoto & Yuji, 2001; Yeh, Ke, Yang, & Meng, 2005). Unsupervised methods do not require training data such as human-made summaries to train the summarizer (Nomoto & Yuji, 2001; Buckley & Walz, 1999; Gong & Liu, 2001).

Recently, many generic document summarization methods using Latent Semantic Analysis (LSA) have been proposed (Gong & Liu, 2001; Li, Li, & Wu, 2006; Yeh et al., 2005; Zha, 2002). The LSA-related methods represent a sentence by means of a linear combination of semantic features. But singular vectors obtained from LSA are not sparse. Hence, the semantic features obtained from LSA are composed of a great number of negative and positive weighted terms. Also, the meaning of the semantic features cannot be captured intuitively, and the scope of their meaning may be obscure. Furthermore, the weights of semantic features that are the elements of a linear combination representing a sentence can have both positive and negative values. A sentence, therefore, must be represented as a linear combination of many less important semantic features. As a result, LSA-related methods of document summarization may fail to extract meaningful sentences (Lee & Seung, 1999; Zha, 2002).

In this paper, we propose a new unsupervised generic document summarization method using Non-negative Matrix Factorization (NMF). The proposed method has the following advantages: First, it is an unsupervised method and does not require training summaries for the summarizer and the training step. Second, the semantic feature vectors extracted from NMF can be interpreted more intuitively than those extracted from LSA-related methods, since the components of the former have only non-negative values and they are very sparse while the components of the latter have both positive and negative values and there are few zero values. Moreover, a sentence can be represented as a linear combination of a few intuitive semantic features. Finally, the scope of meaning of the semantic features is narrow, since they are very sparse and, accordingly, subtopics of a document can be identified more successfully. Therefore, there is greater possibility of extracting important sentences.

The remainder of this paper is organized as follows: Section 2 describes related work regarding generic document summarization and the LSA method; Section 3 describes NMF; Section 4 presents a comparison of summarization methods using NMF and LSA; Section 5 describes NMF summarization; Section 6 describes the performance evaluation. Finally, in Section 7, we conclude the paper with directions for future research.

2. Related work

Kupiec, Pedersen, and Chen (1995) proposed a trainable summarizer, a kind of supervised method. statistical classifier is constructed using Bayes' rule and hand-selected training summaries. Chuang and Yang (2000) proposed a supervised method that uses a machine learning technique to extract sentences. Their method divides sentences into segments that are represented by a set of predefined features. The summarizer is trained to extract the important sentence segments based on this set of features. Amini and Gallinari (2002) proposed a semi-supervised algorithm to train classification models for text summarization. Their method uses classification expectation maximization (CEM) as a semi-supervised learning method, and it makes use of a few labeled data items together with a large amount of unlabeled data for training. Yeh et al. (2005) proposed a new trainable summarizer for document summaries. Their summarizer uses several kinds of document features. Sentence positions are ranked and a genetic algorithm is used to train the score functions. Shen, Sun, Li, Yang, and Chen (2007) proposed a document summarization method using conditional random fields; this is a supervised method that has benefits of an unsupervised method.

Mihalcea (2005) proposed TextRank, which is used for unsupervised extractive summarization. It relies on an iterative graph-based ranking algorithm. Sentences that are highly recommended by other sentences are extracted as a summary. However, it does not identify disjointed subtopics. Nomoto and Yuji (2001) proposed an unsupervised text summarization method. It uses modified X-means to find diverse topic areas in text and a simple sentence weighting model to identify the most important sentence from each topic area.

Kruengkrai and Jaruskulchai (2003) exploited both local and global properties (LGP) of sentences in order to find clusters of significant words in each sentence and calculate a local clustering score. It then calculates a global connectivity score for the global properties of a given sentence by using the cosine similarity between sentences. It subsequently extracts the sentence having the largest combination score with respect to the local clustering score and the global connectivity score.

Xu, Liu, and Gong (2003) proposed two generic document summarization methods. The first method uses a relevance measure (RM). It splits a document into a set of candidate sentences. It then extracts high score sentences from the set by using the RM. The second method uses a latent semantic analysis (LSA) to semantically identify important sentences for summary creations. The sentence that has the largest index value with respect to the important singular vector is extracted, and is thus understood as most important. The singular vectors other than the one corresponding to the largest singular value can have both positive and negative components, making sentence ranks by singular vector component values less meaningful (Zha, 2002).

Zha (2002) proposed generic summarization using sentence clustering and the mutual reinforcement principle (MRP). Their method clusters sentences of documents into several topical groups. Sentences are extracted from each topical group by their saliency scores, which are computed using the MRP; this is a modified LSA method. This method guarantees that the elements of a singular vector with respect to semantic feature values will be only positive values, even though the semantic features do not necessarily identify subtopics. Yeh et al. (2005) proposed a document summarization method using LSA and a text relationship map (TRM). Their method uses LSA to derive the semantic matrix of a document and uses semantic representation of a sentence to construct a semantic text relationship map. TRM is constructed using the similarity between semantic representations, and important sentences are extracted by using the number of links in the TRM. This method does not consider subtopics for document summarization. Li et al. (2006) extended generic multi-document summarization using LSA to query-based document summarization.

2.1. Document summarization using latent semantic analysis

In this paper, we define the matrix notation as follows: Let X^{*}_{j} be the jth column vector of matrix X, X_i^{*} be the ith row vector, and X_{ij} be the element of the ith row and jth column.

The document summarization method using LSA applies singular value decomposition (SVD), delineated in Eq. (1), to summarize documents. This method decomposes matrix A into three matrices, U, D, and V^T (Gong & Liu, 2001).

$$\begin{aligned} A &= UDV^{\mathrm{T}} \\ A &\approx \tilde{U}\tilde{D}\tilde{V} \end{aligned} \tag{1}$$

where A is a $m \times n$ terms-by-sentences matrix, m is the number of terms, and n is the number of sentences in a document. U is an $m \times n$ orthonormal matrix of eigenvectors of AA^T (left singular vectors) and V is an $n \times n$ orthonormal matrix of eigenvectors of A^TA (right singular vectors). $D = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is an $n \times n$ diagonal matrix whose diagonal elements are non-negative eigenvalues sorted in descending order. $\tilde{U}$ is an $m \times r$ matrix, where $\tilde{U}_{ij} = U_{ij}$ if $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant r$. $\tilde{D}$ is an $r \times r$ matrix, where $\tilde{D}_{ij} = D_{ij}$ if $1 \leqslant i,j \leqslant r$. $\tilde{V}$ is an $n \times r$ matrix, where $\tilde{V}_{ij} = V_{ij}$ if $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant r$. And r is very much smaller than n. In the method using LSA, the i'th column vector A^{*}_{i} of matrix A is the weight vector of the ith sentence, and is represented as a linear combination of the left eigenvectors U^{*}_{j}, which are semantic feature vectors, as shown in Eq. (2). That is, the weight of the jth semantic vector U^{*}_{j} corresponding to the sentence vector $A^{*}_{i}$is $\sigma_j V_{ij}$

$$A_{*_i} = \sum_{j=1}^{r} \sigma_j \tilde{V}_{ij} \tilde{U}_{*_j} \tag{2}$$

Starting from the first row of V^T, the sentence corresponding to the column that has the largest index value with the right singular vector is extracted, to be included in summarized sentences (Gong & Liu, 2001).

Example 1. We illustrate an example using LSA algorithm. Let r be 3. LSA decomposes Matrix A into U,D, and V, as shown in Fig. 1a. Fig. 1b shows an example of sentence representation. The column vector A^{*}_{3} corresponding to the third sentence is represented as a linear combination of feature vectors U^{*}_{j} and their weights $\sigma_j V_{ij}$.

$$\begin{array}{ccccccccc}
A & & U & & D & & V^t \\
\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\\10&11&12\end{bmatrix} & \approx & \begin{bmatrix}-0.1409&-0.8247&0.4872\\-0.3439&-0.4263&-0.4630\\-0.5470&-0.0278&-0.5355\\-0.7501&0.3706&0.5113\end{bmatrix} & \times & \begin{bmatrix}25.4624&0&0\\0&1.2907&0\\0&0&0\end{bmatrix} & \times & \begin{bmatrix}-0.5045&-0.5746&-0.6445\\0.7608&0.0571&-0.6465\\0.4082&-0.8165&0.4082\end{bmatrix} \\
A_{*3} & & U_{*1} \quad U_{*2} & & & &
\end{array}$$

(a) result of LSA decomposition

$$\begin{array}{ccc}
\begin{bmatrix}3\\6\\9\\12\end{bmatrix} & \approx 25.4624 \times (-0.6445) \times \begin{bmatrix}-0.1409\\-0.3439\\-0.5470\\-0.7501\end{bmatrix} & + \ 1.2907 \times (-0.6465) \times \begin{bmatrix}-0.8247\\-0.4263\\-0.0278\\0.3706\end{bmatrix} \\
A_{*3} & \mathrm{D}_{11} \quad \mathrm{V}_{13} \quad \mathrm{U}_{*1} & \mathrm{D}_{22} \quad \mathrm{V}_{23} \quad \mathrm{U}_{*2}
\end{array}$$

(b) example of sentence representation using semantic features and semantic variables

Fig. 1. Example of sentence representation using semantic features and their weights.

3. Non-negative matrix factorization

Unlike LSA, NMF represents individual objects as a non-negative linear combination of part information extracted from a large volume of object. Studies on human cognition show that we use only the summation of non-negative data when we recognize an object as a combination of partial information (Lee & Seung, 1999). This method can deal with a large volume of information efficiently, since the original non-negative matrix is decomposed into a sparsely distributed representation of two non-negative matrices (Lee & Seung, 1999; Xu, Liu, & Gong, 2003).

NMF is employed to decompose a given $m \times n$ non-negative matrix A into a multiplication of an $m \times r$ non-negative semantic feature matrix (NSFM), W, and an $r \times n$ non-negative semantic variable matrix (NSVM), H, as shown in Eq. (3)

$$A \approx WH \tag{3}$$

where r is usually chosen to be smaller than m or n so that the total sizes of W and H are smaller than that of matrix A.

We use the Frobenius norm as an objective function to satisfy the approximation condition $\widetilde{A} = WH$. The Frobenius norm is shown in Eq. (4) (Lee & Seung, 1999, 2001; Wild, Curry, & Dougherty, 2003):

$$\Theta_E(W,H) \equiv \|A - WH\|_F^2 \equiv \sum_{j=1}^{m} \sum_{i=1}^{n} \left(X_{ji} - \sum_{l=1}^{r} W_{jl} H_{li} \right)^2 \tag{4}$$

This is lower bounded by zero, and clearly vanished if and only if $A = WH$. W and H are continuously updated until $\Theta_E(W,H)$ converges under the predefined threshold or exceeds the number of repetitions. The update rules are as follows:

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} \frac{(W^T A)_{\alpha\mu}}{(W^T W H)_{\alpha\mu}} \qquad W_{i\alpha} \leftarrow W_{i\alpha} \frac{(AH^T)_{i\alpha}}{(WHH^T)_{i\alpha}} \tag{5}$$

A column vector corresponding to the j'th sentence, A^*_j, can be represented as a linear combination of the semantic feature vectors W^*_l and the semantic variable H_{lj} as follows:

$$A_{*j} = \sum_{l=1}^{r} H_{lj} W_{*l} \tag{6}$$

Example 2. We illustrate the example using the NMF algorithm: Let r be 2, the number of repetitions 50, and the tolerance 0.001. When the initial elements of the W and H matrices are 0.5, the non-negative matrix A is decomposed into two non-negative matrices, W and H, as shown in Fig. 2a. Fig. 2b shows an example of sentence representation using NMF. The column vector A^*_3 corresponding to the third sentence is represented as a linear combination of the semantic feature vectors W^*_l and the semantic variable column vector H^*_3.

In contrast to LSA, NMF decomposes a sparse matrix into two sparse matrices. Fig. 3 shows this property of the NMF. Here, a non-zero ratio for a matrix means the value of the number of non-zero elements divided by the total number of elements of the matrix. The non-negative matrix A is a n-by-n matrix, and the values of n are set to 100, 200, 300, and 400. Non-zero entries are chosen at random. The number of semantic features, r, is chosen as 10% of n. The non-zero ratios of A are chosen as 0.5%, 1%, 2%, 3%, 5%, 7%, 10%, 30%, 60%, and 99%. The matrices W and H are obtained by using NMF. The matrices U and V are obtained by using LSA.

$$\underset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}} \approx \underset{W}{\begin{bmatrix} 0.1487 & 1.5998 \\ 0.6610 & 0.9676 \\ 1.1481 & 0.5727 \\ 1.6129 & 0.4066 \end{bmatrix}} \times \underset{H}{\begin{bmatrix} 6.1136 & 6.6784 & 7.1784 \\ 0.0854 & 0.5932 & 1.2245 \end{bmatrix}} = \underset{\widetilde{A}}{\begin{bmatrix} 1.0457 & 1.9420 & 3.0263 \\ 4.1237 & 4.9813 & 5.9297 \\ 7.0679 & 8.0071 & 8.9470 \\ 9.8953 & 11.0127 & 12.0759 \end{bmatrix}}$$

A_{*3} W_{*1} W_{*2} H_{*3}

(a) result of NMF decomposition

$$\underset{A_{*3}}{\begin{bmatrix} 3 \\ 6 \\ 9 \\ 12 \end{bmatrix}} \approx \underset{\mathrm{H}_{13}}{7.1784} \times \underset{\mathrm{W}_{*1}}{\begin{bmatrix} 0.1487 \\ 0.6610 \\ 1.1481 \\ 1.6129 \end{bmatrix}} + \underset{\mathrm{H}_{23}}{1.2245} \times \underset{\mathrm{W}_{*2}}{\begin{bmatrix} 1.5998 \\ 0.9676 \\ 0.5727 \\ 0.4066 \end{bmatrix}}$$

(b) example of sentence representation using semantic features and semantic variables

Fig. 2. Example of NMF algorithm.

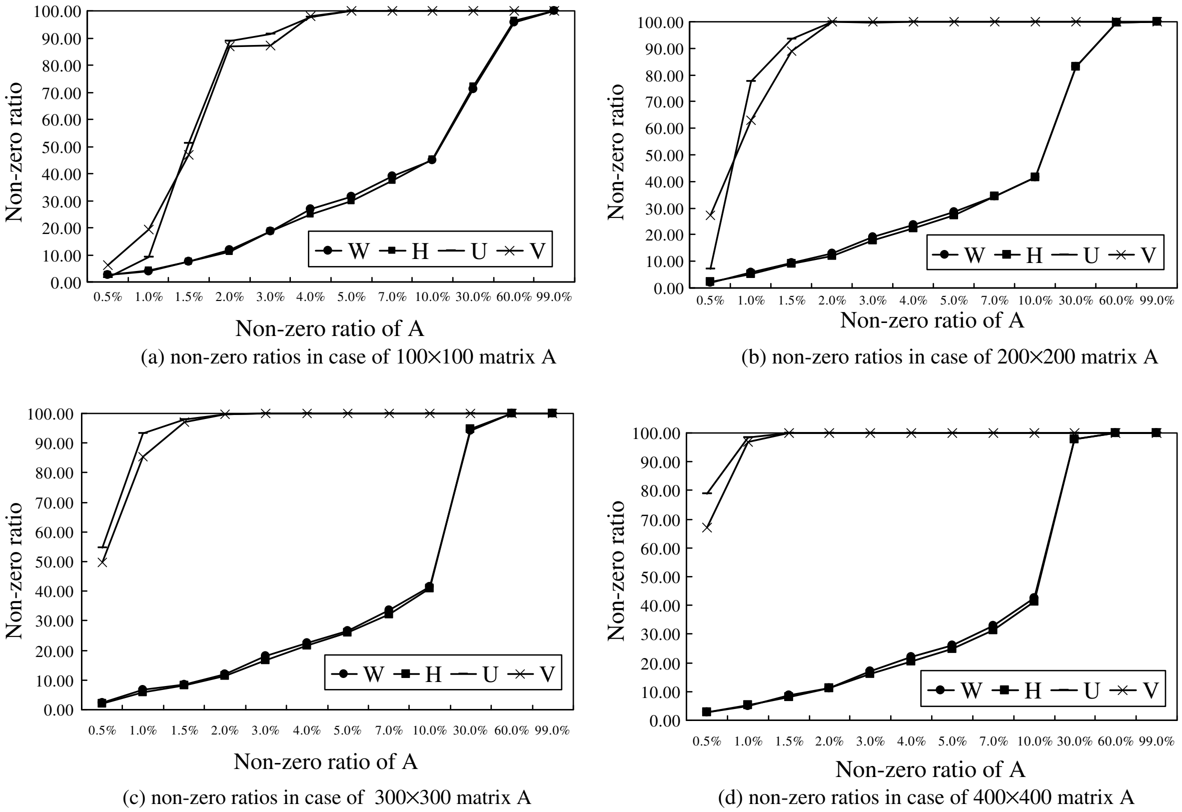


Fig. 3. Comparison of non-zero ratios between NMF and LSA decompositions.

As shown in Fig. 3, NMF decomposition produces sparse matrices while LSA produces non-sparse matrices.

4. Comparison of LSA and NMF as summarization methods

Table 1 shows some extracted sentences from 20 documents related to the topic "Tourism in Great Britain" (Hoa, 2005). Table 2 shows the terms-by-sentences matrix A obtained by preprocessing a set of sentences in Table 1. Matrix A is composed of 396 terms and 57 sentences. Tables 3 and 4 illustrate the cases of applying LSA and NMF to matrix A.

As shown in Table 2, the terms-by-sentences matrix A is very sparse. Table 3 shows 10 semantic feature vectors $U^*_1 \ldots, U^*_{10}$ obtained from SVD decomposition of matrix A, the weight values $\sigma_1 V_{20,1}, \ldots, \sigma_{10} V_{20,10}$ of semantic feature vectors with respect to the sentence S20, the original sentence vector, the sentence vector calculated from the weight values, and the semantic feature vectors.

Table 1
Some sentences related with the topic "Tourism in Great Britain"

The number of sentence	Sentences
S1	TOURIST arrivals to the UK in 1991 are forecast to recover sharply after the steep decline earlier this year caused by the Gulf war. The British Tourist Authority said incoming tourist numbers had already increased significantly after falling 18 percent in the first two months of this year from the levels of the corresponding period of 1990
S2	The increases were achieved in spite of a fall in the number of North American visitors Visits by North Americans fell 6 per cent to 600,000 in the first quarter However, the number of visitors from western Europe rose 12 per cent to 23 m – higher than in any previous first quarter. A RECORD 185 m tourists visited Britain in the 12 months to March, 8 percent more than the previous year – and the British Tourist Authority said yesterday that it was expecting even higher numbers this year
...	...
S20	The increases were achieved in spite of a fall in the number of North American visitors Visits by North Americans fell 6 per cent to 600,000 in the first quarter However, the number of visitors from western Europe rose 12 per cent to 23 m – higher than in any previous first quarter. A RECORD 185 m tourists visited Britain in the 12 months to March, 8% more than the previous year – and the British Tourist Authority said yesterday that it was expecting even higher numbers this year
...	...

Table 2
Terms-by-sentences matrix A

Term		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	...	S20	...	S57
1	Tourist	3	2	0	2	1	0	0	0	0	0	...	2	...	1
2	Arrival	1	0	0	0	0	0	0	0	0	0		0		0
3	UK	1	1	0	0	1	0	0	0	0	0	...	0	...	1
4	Forecast	1	0	0	0	0	0	0	0	0	0	...	0	...	0
5	Recover	1	0	0	0	0	0	0	0	0	0	...	0	...	0
6	Sharply	1	0	0	0	0	0	0	0	0	0	...	0	...	0
7	Steep	1	0	0	0	0	0	0	0	0	0	...	0	...	0
8	Decline	1	0	0	0	0	0	0	0	0	0	...	0	...	0
9	Earlier	1	0	0	0	0	0	0	0	0	0	...	0	...	0
10	Year	2	2	1	0	0	1	0	0	0	0	...	2	...	0
11	Cause	1	0	0	0	0	0	0	0	0	0	...	0	...	0
12	Gulf	1	0	0	0	0	0	0	0	0	0	...	0	...	0
13	War	1	0	0	0	0	0	0	0	0	0	...	0	...	0
14	British	1	2	0	1	1	0	0	0	0	0	...	2	...	0
15	Authority	1	1	0	1	1	0	0	0	0	0	...	1	...	1
16	Income	1	0	0	0	0	0	0	0	0	0	...	0	...	0
17	Increase	1	1	0	0	0	0	0	0	0	0	...	1	...	0
18	Significantly	1	0	0	0	0	0	0	0	0	0	...	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
396	Return	0	0	0	0	0	0	0	0	1	0	...	0	...	0

Table 3
Representation of sentences by means of LSA, Eq. (2)

Term		Semantic feature										Sentence S20	
		U^*_1	U^*_2	U^*_3	U^*_4	U^*_5	U^*_6	U^*_7	U^*_8	U^*_9	U^*_{10}	Original	$\sum_{j=1}^{10}\sigma_j V_{20j} U_{*j}$
...	...	...	...	...	...	...	...	...	...	...	...	...	...
13	War	−0.04	0.03	−0.004	0.01	0.09	−0.09	0.03	0.05	−0.01	0.12	0	0.00
14	British	−0.11	−0.02	−0.05	0.09	−0.06	0.03	−0.13	0.11	0.18	−0.01	1	1.00
15	Authority	−0.17	−0.003	−0.17	0.27	−0.07	0.06	0.05	−0.01	0.03	−0.23	1	1.00
16	Income	−0.02	−0.01	0.02	0.01	−0.02	−0.05	−0.06	−0.07	−0.02	0.02	0	0.00
17	Increase	−0.11	−0.01	0.02	−0.04	0.12	0.06	0.14	0.07	−0.11	0.18	1	1.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...
396	Return	−0.01	−0.004	0.01	0.007	0.04	−0.03	0.004	0.02	0.01	0.031	0	0.00
Weight $\sigma_j V_{20j}$		−2.96	1.77	−0.04	0.22	0.02	−0.84	1.07	−0.56	1.86	1.67		

Table 4
Representation of sentence by means of NMF, Eq. (8)

Term		Semantic feature										Sentence S20	
		W^*_1	W^*_2	W^*_3	W^*_4	W^*_5	W^*_6	W^*_7	W^*_8	W^*_9	W^*_{10}	Original	$\sum_{l=1}^{10} H_{l20} W_{*l}$
...	...	...	...	...	...	...	...	...	...	...	...	...	...
13	War	0	0	0	0	1.12	0.46	0.42	0.17	0	0.04	0	0.11
14	British	0	0.68	0.44	2.11	0.30	0	0	0	0.41	0.13	1	0.89
15	authority	0	0.60	0	2.17	0	1.10	0	0.21	2.77	0	1	1.05
16	income	0	0	0.35	0.27	0.31	0	0	0.00	0	0.03	0	0.11
17	increase	0.07	0	0	0.26	0.67	0.77	1.26	1.38	0	0.76	1	1.01
...	...	...	...	...	...	...	...	...	...	...	...	...	...
396	Return	0	0	0	0	0	0	0.09	0.10	0	0.08	0	0.07
weight H_{j20}		0	0.07	0	0.40	0	0.01	0	0.65	0	0		

The method using LSA indicates that the first semantic feature vector U^*_1 has the largest weight values, since the first column value of V is multiplied by the largest eigenvalue. Therefore, we extract the sentence having the largest value among weights of U^*_1 to summarize documents.

Table 4 presents 10 semantic feature vectors $W^*_1,\ldots,W^*_{10}$ obtained from NMF decomposition of matrix A, the weight values $H_{1,20}\ldots,H_{10,20}$ of semantic feature vectors with respect to the sentence S20, the original sentence vector, the sentence vector calculated from the weight values, and the semantic feature vectors.

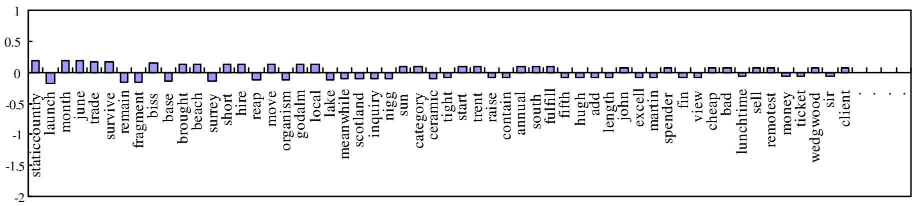
Comparing the semantic feature vectors in Tables 3 and 4, it is seen that there are many negative values and few zero values in Table 3 whereas there are no negative values and many zero values in Table 4. That is, the semantic feature vectors obtained by using the NMF are sparser than those obtained using LSA. Thus, the scope of the meaning of the semantic features obtained by using the NMF is clearer and narrower than that obtained by using LSA. This indicates that the method using NMF provides better performance in identifying subtopics of a document, as compared with the methods using LSA. Furthermore, the semantic feature vectors in Table 4 intuitively make more sense than those in Table 3, because the former have non-negative values and the latter have both negative and positive values. A more intuitive explanation is provided through Examples 3 and 4.

The weight values of semantic variable vectors (H_{j20}) obtained using NMF are also sparser than those ($\sigma_j V_{20j}$) obtained using LSA. This indicates that the methods using LSA represent a sentence as a linear combination of many non-intuitive and less important semantic feature vectors, whereas the method using NMF represents a sentence as a linear combination of only a few intuitive and directly related-semantic feature vectors.

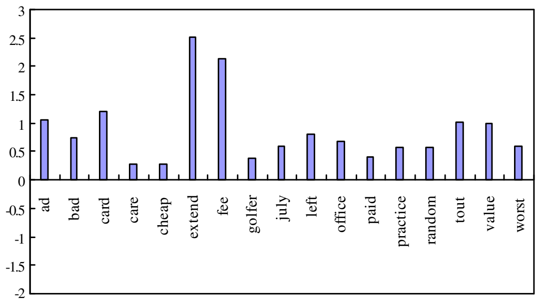
The method using LSA extracts the sentence from the first row of V corresponding to the semantic feature vector having the largest weight, and extracts the next sentence from the second row V corresponding to the semantic feature vector having the next largest weight. However, the method using NMF extracts the sentence having the largest weight with respect to how much the sentence reflects major topics, which are represented as semantic features. Therefore, the method using NMF has a greater likelihood of extracting semantically important sentences, which are closer to the major topics, as compared with the method using LSA.

Example 3. We analyzed examples of semantic features using LSA and NMF. Fig. 4 shows the term weights included in the semantic features U^*_1 in Table 3 (LSA) and W^*_1 in Table 4 (NMF), respectively. The total number of terms is 396. The number of non-zero weights of the semantic feature vectors from LSA and NMF are 385 and 52, respectively. The average term weights are 0.035 and 0.042, respectively. The average values are nearly the same. Here, we take the absolute values for U^*_1. However, the maximum value of the term weights in the semantic feature vector from NMF (2.52) is much larger than that from LSA (0.187). The number of non-zero values greater than a tenth of the maximum values of NMF (17) is considerably smaller that of LSA (231). The standard deviation for term weights from NMF (0.217) is also much greater than that of LSA (0.036). These results indicate that the semantic features from NMF consist of very few terms that have important meanings. In other words, NMF can more intuitively find comprehensible semantic features. These semantic features can be used more appropriately for subtopics of target documents.

Example 4. Fig. 5 shows an example of NMF and LSA representations for a CNN news story. We randomly selected 10 CNN political news stories. The number of semantic features, r, is chosen as 10. One CNN news story is titled, "A'very personal victory' for McCain in New Hampshire". The NMF representation of this news contains 4 semantic features (W^*_3, W^*_5, W^*_7, W^*_8). The news is represented as a linear combination of these features. The values of H are the weights of the semantic features. The terms in the news are also shown in semantic features. Each semantic feature has a few terms. Their weights are all non-negative. Each semantic feature includes partial information of the news. With this information, we can roughly grasp the meaning of the semantic features. In contrast, the semantic features from the LSA representation have almost all terms found in the 10 CNN news articles whose values are positive or negative. Most of the terms in semantic features from LSA do not exist in the stories, and are eliminated when the semantic features are summed up to form the story. Therefore, it is not possible to grasp the meaning of the semantic features obtained from LSA representation through partial information of the news.



(a) Weights of terms in the semantic feature (U_{*1}) from LSA



(b) Weights of terms in the semantic feature (W_{*1}) from NMF

Fig. 4. Comparison of term weights of the semantic features from LSA, NMF.

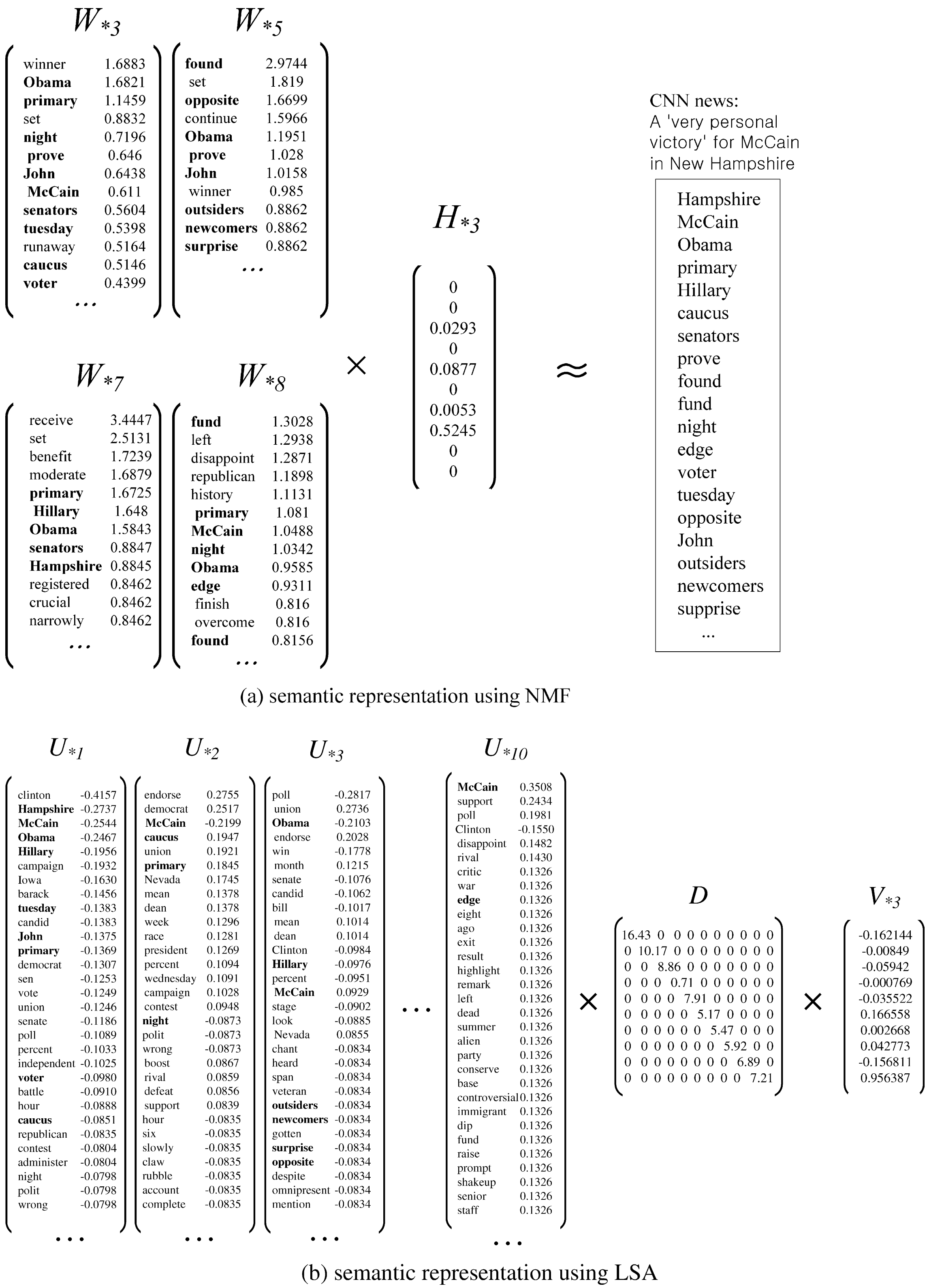


Fig. 5. Semantic representation of the CNN news story "A 'very personal victory' for McCain in New Hampshire".

5. Generating generic summaries

In this section, we propose a method to create generic document summaries by selecting sentences using NMF. The proposed method consists of a preprocessing step and a summarization step. In the following subsections, the two steps are described in detail.

We give a full explanation of the two phases in Fig 6.

5.1. Preprocessing

In the preprocessing step of generating generic document summaries, after a given English document is decomposed into individual sentences, all stopwords are removed by using Rijsbergen's stopwords list and word stemming is performed by Porter's stemming algorithm (Frankes & Baeza-Yates, 1992). A term-frequency vector for each sentence in the document is then constructed by Eq. (7).

Let element A_{ji} be the weighted term-frequency of term j in sentence i in a $m \times n$ terms-by-sentences matrix A.

$$A_{ji} = Wgt(j, i) \tag{7}$$

where $Wgt(j, i)$ is the weight for term j in sentence i. Several weighting schemes are when constructing the weighted term-frequency vector A^{*}_{i}. A weighting scheme measures the importance of a term not in a sentence but in the overall document. Several areas in information retrieval use weighting schemes, as they enhance performance in most cases. (Frankes & Baeza-Yates, 1992). In Section 6.3, we evaluate several weighting schemes and explain how these weighting schemes affect the summarization results.

5.2. Generic document summarization by non-negative matrix factorization

The summarization step selects sentences for generic summarization by using NMF. We perform NMF on A to obtain the non-negative semantic feature matrix W while the non-negative semantic variable matrix H is obtained using Eqs. (4) and (5).

We propose a novel method to select sentences based on NMF and define the *Generic Relevance of a Sentence (GRS)* as follows:

$$\text{Generic Relevance of a } j\text{th sentence} = \sum_{i=1}^{r} (H_{ij} \cdot \text{weight } (H_{i*})) \tag{8}$$

$$\text{weight } (H_{i*}) = \frac{\sum_{q=1}^{n} H_{iq}}{\sum_{p=1}^{r} \sum_{q=1}^{n} H_{pq}} \tag{9}$$

The weight (H_i^{*}) is the relative relevance of the i'th semantic feature (W^{*}_{i}) among all semantic features. The generic relevance of a sentence refers to how much the sentence reflects major topics, which are represented as semantic features.

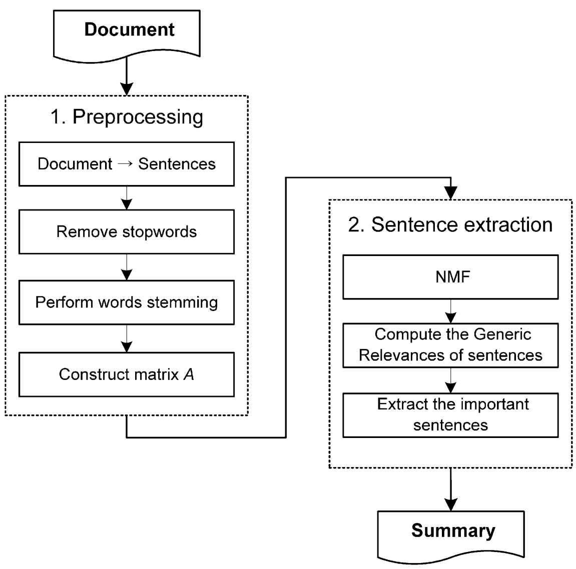


Fig. 6. Generic summarization method using NMF.

Table 5
The four sentences

The number of sentences	*S1*	Ordinary Differential Equations
	S2	Oscillation Theory for Neutral Equations
	S3	Oscillation of Delay Differential of Equations
	S4	Sync Methods for Quadrature and Differential Equations

5.3. The proposed generic document summarization algorithm

The proposed algorithm for generic document summarization is as follows:

1. Decompose the document *D* into individual sentences, and let *k* be the number of sentences for generic document summarization.
2. Perform stopwords removal and word-stemming operations.
3. Construct the terms-by-sentences matrix *A*.
4. Perform NMF on matrix *A* to obtain matrix *H*.
5. For each sentence, calculate its generic relevance.
6. Select *k* sentences with the highest generic relevance values.

Example 5. We illustrate an example of sentence extraction using the proposed method as follows: Table 5 shows four sentences. We generate matrix *A* by preprocessing a set of sentences in Table 5 and decomposing matrix *A* into a semantic feature matrix and a semantic variable matrix using NMF. Fig. 7 illustrates the sentence extraction process from the set of sentences in Table 5. We compute the *GRS* and then extract the sentence*S2* corresponding to the semantic variable column vector H^{*}_{2} having the largest *GRS* value (0.53).

6. Experiments

6.1. Data set and evaluation system

We used the DUC2006 data set as test documents. The Document Understanding Conference (DUC[1]) is an international conference for performance evaluations of proposed system by comparing manual summaries by experts with summaries of the proposed system. The systems participating in DUC2006 are constrained to query-based multi-document summarization. However, our proposed system is based on generic summarization. Therefore, we implemented 4 kinds of generic summarization methods (RM, LSA, MRP, LGP) using the DUC2006 data set. To compare the performances, we used the ROUGE evaluation software package, which compares various summary results from several summarization methods with summaries generated by humans.

Fig. 8 illustrates how the evaluations in this experiment are performed. As test data, we randomly selected 50 documents from the DUC2006 data set. Each document has a human-produced summary. Our methods (NMF) and 4 other methods produce summaries using test documents. These summaries are input to ROUGE software to yield the ROUGE evaluation values.

We use ROUGE (recall-oriented understudy for gisting evaluation) to evaluate the proposed method. ROUGE has been applied by DUC for performance evaluation (Lin, 2004). As experimental data, we use the testing data from DUC 2006. This testing data consists of 50 topics and 25 documents related to each topic (Hoa, 2005).

6.2. Performance evaluation measure

We conducted a performance evaluation of the document summarization methods using 50 given topics from DUC 2006.

ROUGE includes five automatic evaluation methods, *ROUGE-N*, *ROUGE-L*, *ROUGE-W*, *ROUGE-S*, and *ROUGE-SU* (Lin, 2004). Each method estimates recall, precision, and *f*-measure between human written reference summaries and candidate summaries of the proposed system. *ROUGE-N* uses *n*-gram recall between a candidate summary and a set of reference summaries. *ROUGE-N* is computed as follows:

$$ROUGE-N = \frac{\sum_{S\in\{\text{ReferenceSummaries}\}}\sum_{gram_n\in S} Count_{\text{match}}(gram_n)}{\sum_{S\in\{\text{ReferenceSummaries}\}}\sum_{gram_n\in S} Count\ (gram_n)} \tag{10}$$

where *n* is the length of the *n-gram*, gram_n, and $Count_{match}(gram_n)$ is the maximum number of *n-grams* co-occurring in a candidate summary and a set of reference summaries (Lin, 2004). *ROUGE-L* computes the ratio between the length of the summaries' longest common subsequence (LCS) and the length of the reference summary, as delineated by Eq. (11):

[1] http://www-nlpir.nist.gov/projects/duc/index.html

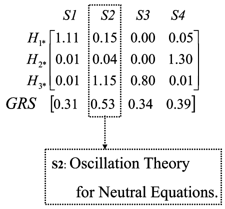


Fig. 7. Sentence extraction using the *GRS*.

Table 6
Results of *t*-test between RM and NMF

Rouge		μ_0 RM	NMF			*t*-test result
			Mean	Stderr	*t*	
Recall	1	0.2298	0.2763	0.008143	5.7169	Accept
	L	0.2108	0.2541	0.007446	5.8105	Accept
	W	0.0634	0.0732	0.002109	4.6055	Accept
	SU	0.0666	0.0853	0.003032	6.1885	Accept
Precision	1	0.3387	0.3567	0.006481	2.7848	Accept
	L	0.2958	0.3291	0.006585	5.0585	Accept
	W	0.1574	0.1773	0.003809	5.2222	Accept
	SU	0.1032	0.1111	0.003448	2.3080	Accept
F-measure	1	0.2686	0.3070	0.007048	5.4407	Accept
	L	0.2469	0.2825	0.006480	5.4991	Accept
	W	0.0899	0.1023	0.002524	4.9230	Accept
	SU	0.0823	0.0952	0.003083	4.1609	Accept

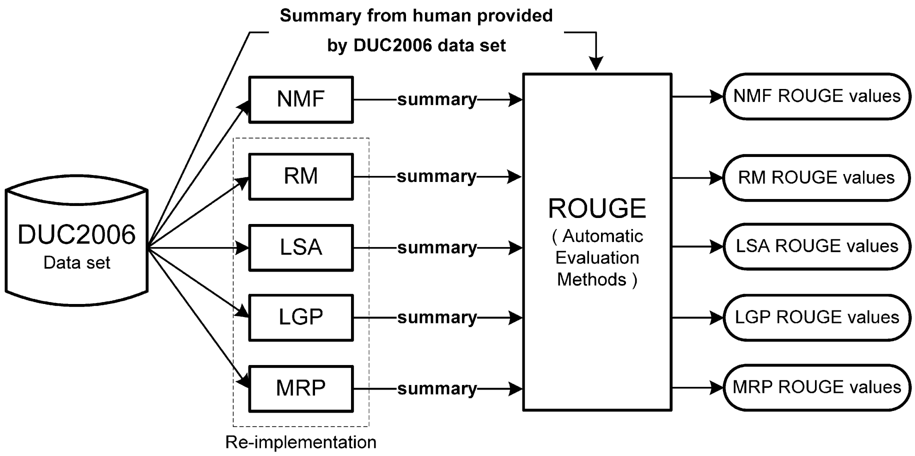


Fig. 8. Evaluation system.

$$R_{lcs} = \frac{LCS(X,Y)}{m}, P_{lcs} = \frac{LCS(X,Y)}{n}, F_{lcs} = \frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs}+\beta^2 P_{lcs}} \tag{11}$$

where m is the length of reference summary sentence X and n is the length of candidate sentence Y. $LCS(X,Y)$ is the length of a LCS of X and Y. R_{lcs} is a recall of $LCS(X,Y)$, P_{lcs} is the precision of $LCS(X,Y)$, and $\beta = P_{lcs}/R_{lcs}$ (Lin, 2004). *ROUGE-W* uses the weighted LCS that favors LCS with consecutive matches. *ROUGE-S* uses the overlap ratio of the skip-bigram between a candidate summary and a set of reference summaries, as given by Eq. (12):

$$R_{skip2} = \frac{SKIP2(X,Y)}{C(m,2)}, P_{skip2} = \frac{SKIP2(X,Y)}{C(n,2)}, F_{skip2} = \frac{(1+\beta^2)R_{skip2}P_{skip2}}{R_{skip2}+\beta^2 P_{skip2}} \tag{12}$$

where $SKIP2(X,Y)$ is the number of skip-bigrams between X and Y, and β is the relative importance of P_{skip2} and R_{skip2}, P_{skip2} being the precision of $SKIP2(X,Y)$ and R_{skip2} a recall of $SKIP2(X,Y)$. $C(\ )$ is a combination function (Lin, 2004). *ROUGE-SU* is an extension of *ROUGE-S* with the addition of unigram as the counting unit (Lin, 2004).

6.3. Weighting schemes

We evaluated the summarization performances for the eight weighting schemes (Frankes & Baeza-Yates, 1992; Baeza-Yates & Ribeiro-Neto, 1999; Gong & Liu, 2001). The equation of the weighting schemes is the same as that from the SMART system (Gong & Liu, 2001; Buckley & Walz, 1999):

$$\text{No weight} \quad Wgt(j,i) = t_{ji} \tag{13}$$

$$\text{Ordinary weight} \quad Wgt(j,i) = t_{ji} \cdot \log(N/n(i)) \tag{14}$$

$$\text{Binary weight} \quad Wgt(j,i) = 1 \text{ if term } i \text{ appears at least once in the sentence; otherwise } Wgt(j,i) = 0 \tag{15}$$

$$\text{Modified binary weight} \quad Wgt(j,i) = \log(N/n(i)) \text{ ifterm } i \text{ appears at least once in the sentence; otherwise } Wgt(j,i) = 0 \tag{16}$$

$$\text{Augmented weight} \quad Wgt(j,i) = 0.5 + 0.5 \cdot (t_{ji}/\max(t_{ji})) \tag{17}$$

$$\text{Ordinary augmented weight} \quad Wgt(j,i) = (0.5 + 0.5 \cdot (t_{ji}/\max(t_{ji})) \cdot \log(N/n(i)) \tag{18}$$

$$\text{Logarithm weight} \quad Wgt(j,i) = \log(1 + t_{ji}) \tag{19}$$

$$\text{Ordinary logarithm weight} \quad Wgt(j,i) = \log(1 + t_{ji}) \cdot \log(N/n(i)) \tag{20}$$

where $\max(t_{ji})$ is the frequency of the most frequently occurring term in sentence, N is the total number of sentences in the document, and $n(i)$ is the number of sentences that contain term i.

6.4. Results and discussion

In this paper, we conducted two experiments to evaluate the performance of the proposed method. In the first experiment, we conducted a performance evaluation (t-test) using the ROUGE measure with respect to five document summarization methods. For the t-test evaluation, we established several hypotheses. Examples of our hypotheses are "our proposed method (NMF) is superior to RM in ROUGE-1 recall", "our proposed method (NMF) is superior to LSA in ROUGE-W F-measure", etc. The significance level is 5% and the number of samples is 50. The acceptance region of t is $t > t_{0.05}(50) \approx 1.6525$. If t is larger than 1.6525, the hypothesis is accepted; otherwise, it is rejected. In the second experiment, we evaluate the influences of different weighting schemes on the proposed method.

Experiment 1. We evaluated the five different summarization methods: RM, LSA, MRP, LGP, and NMF. In Table 6, RM denotes Gong and Liu's (2001) method using the relevance measure. In Table 7, LSA denotes their method using latent semantic analysis. In Table 8, MRP denotes Zha's (2002) method using the mutual reinforcement principle and sentence clustering. In Table 9, LGP denotes Kruengkrai and Jaruskulchai's (2003) method using the local and global properties of sentences. NMF denotes the proposed generic document summarization algorithm using NMF. The t-test evaluation results are presented in Tables 6–9.

In this experiment, most hypotheses were accepted except a few hypotheses regarding precision. However, the F-measure is more important, because it is the synthesis of recall and precision. Most F-measure tests were accepted. RM showed the poorest performance, because it only uses the cosine similarity between sentences. LGP, however, showed better performance than LSA, because it reflects the local and global properties in documents while LSA cannot consider subtopics successfully. MRP showed better performance than LGP, because it can select sentences that contain major topics by using

Table 7
Results of t-test between LSA and NMF

Rouge		μ_0 LSA	NMF			t-test result
			Mean	Stderr	t	
Recall	1	0.2329	0.2763	0.008143	5.3316	Accept
	L	0.2143	0.2541	0.007446	5.3356	Accept
	W	0.0625	0.0732	0.002109	5.0613	Accept
	SU	0.0702	0.0853	0.003032	5.0040	Accept
Precision	1	0.3428	0.3567	0.006481	2.1483	Accept
	L	0.3098	0.3291	0.006585	2.9312	Accept
	W	0.1699	0.1773	0.003809	1.9260	Accept
	SU	0.1000	0.1111	0.003448	3.2299	Accept
F-measure	1	0.2729	0.3070	0.007048	4.8331	Accept
	L	0.2512	0.2825	0.006480	4.8287	Accept
	W	0.0905	0.1023	0.002524	4.7016	Accept
	SU	0.0819	0.0952	0.003083	4.2984	Accept

Table 8
Results of t-test between MRP and NMF

Rouge		μ_0 MRP	NMF			t-test result
			Mean	Stderr	t	
Recall	1	0.2446	0.2763	0.008143	3.9011	Accept
	L	0.2257	0.2541	0.007446	3.8081	Accept
	W	0.0662	0.0732	0.002109	3.2933	Accept
	SU	0.0796	0.0853	0.003032	1.8927	Accept
Precision	1	0.3429	0.3567	0.006481	2.1288	Accept
	L	0.3387	0.3291	0.006585	−1.4503	Reject
	W	0.1756	0.1773	0.003809	0.4464	Reject
	SU	0.1161	0.1111	0.003448	−1.4313	Reject
F-measure	1	0.2908	0.3070	0.007048	2.2964	Accept
	L	0.2682	0.2825	0.006480	2.2067	Accept
	W	0.0969	0.1023	0.002524	2.1409	Accept
	SU	0.0948	0.0952	0.003083	0.1218	Reject

Table 9
Results of t-test LGP and NMF

Rouge		μ_0 LGP	NMF			t-test result
			Mean	Stderr	t	
Recall	1	0.2445	0.2763	0.008143	3.9141	Accept
	L	0.2266	0.2541	0.007446	3.6961	Accept
	W	0.0657	0.0732	0.002109	3.5497	Accept
	SU	0.0738	0.0853	0.003032	3.8177	Accept
Precision	1	0.3470	0.3567	0.006481	1.5026	Reject
	L	0.3219	0.3291	0.006585	1.0944	Reject
	W	0.1699	0.1773	0.003809	1.9345	Accept
	SU	0.1037	0.1111	0.003448	2.1641	Accept
F-measure	1	0.2848	0.3070	0.007048	3.1397	Accept
	L	0.2641	0.2825	0.006480	2.8511	Accept
	W	0.0949	0.1023	0.002524	2.9635	Accept
	SU	0.0862	0.0952	0.003083	2.8913	Accept

the most important singular vector. NMF showed the best performance, because it uses semantic features that are more intuitively interpretable than those of any of the aforementioned document summarization methods. NMF's use of semantic features allows it to identify subtopics of documents more successfully than the LSA-related methods.

Experiment 2. We compared the influences of eight different weighting schemes described in Section 6.3 for the proposed method. Fig. 9 shows the *ROUGE-1* comparison results for the eight weighting schemes on the proposed method. The*ROUGE-L* comparison results, the *ROUGE-W* comparison results, and the *ROUGE-SU* comparison results are shown in Figs. 10–12, respectively. In this experiment, the No weight for the recall evaluation results showed the best performance among the measures of *ROUGE-1*, *ROUGE-L*, and *ROUGE-W* whereas the Ordinary weight showed the best performance in the measure of

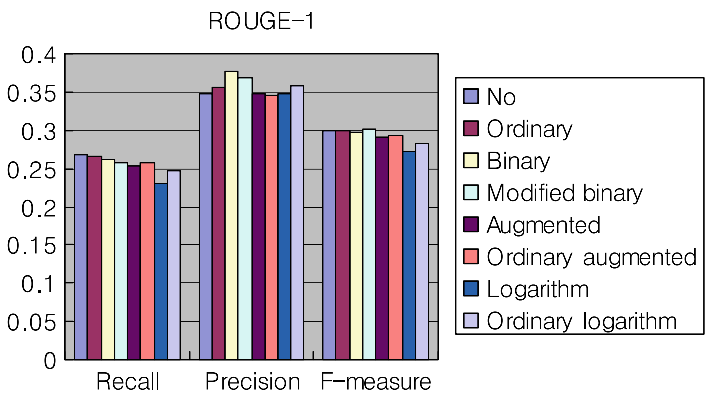


Fig. 9. Comparison using weighting schemes of *ROUGE-1*.

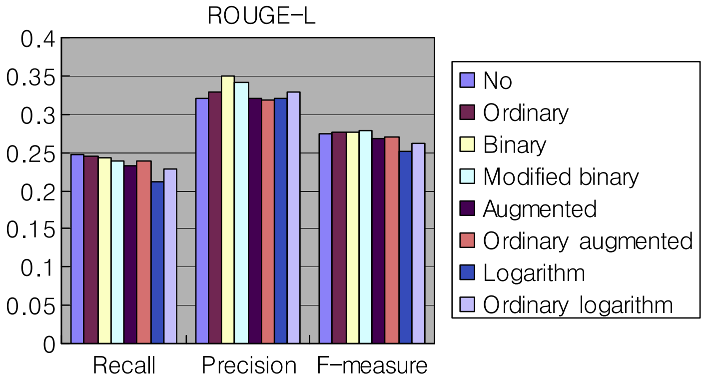


Fig. 10. Comparison using weighting schemes of *ROUGE-L*.

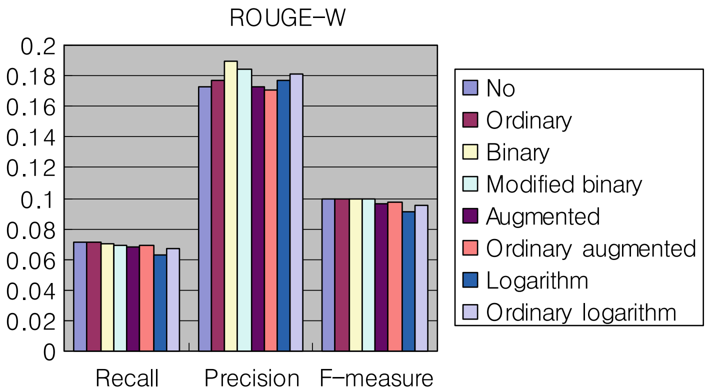


Fig. 11. Comparison using weighting schemes of *ROUGE-W*.

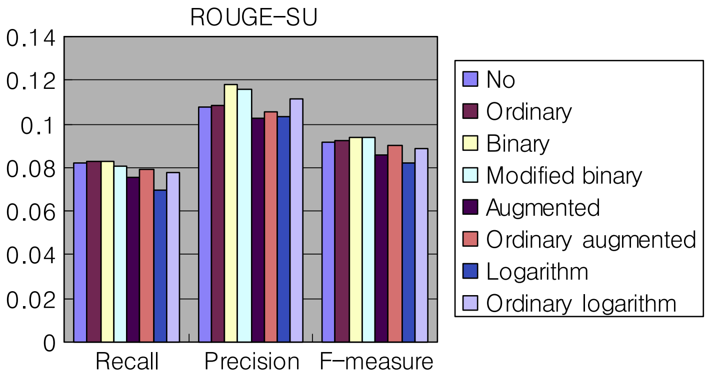


Fig. 12. Comparison using weighting schemes of *ROUGE-SU*.

ROUGE-SU. In the precision evaluation results, the Binary weight showed the best performance among all ROUGE measures. In the F-measure evaluation results, the Modified binary showed the best performance among all ROUGE measures.

7. Conclusions and future research

This paper presents a novel generic document summarization method using the *generic relevance of a sentence* based on NMF. The proposed method has the following advantages: NMF selects more meaningful sentences than the LSA-related methods, because it can use more intuitively interpretable semantic features and is better at grasping the innate structure of documents. As such, it provided superior representation of the subtopics of documents. In Experiment 1, the proposed technique achieved significant improvement over RM, LSA, MRP, and LPG in terms of recall, precision, and F-measure. In Experiment 2, the No weight showed the highest performance in the recall evaluation, the Binary weight showed the highest performance in the precision evaluation, and the Modified binary weight showed the highest performance in the F-measure evaluation.

In future research, automatic relevance feedback and pseudo relevance feedback will be considered in order to enhance the methodology. Furthermore, we plan to investigate the factors that determine which terms are more important than others in summarization.

Acknowledgement

This work was supported by Inha University Research Grant.

References

Amini, M. R., & Gallinari, P. (2002). The use of unlabeled data to improve supervised learning for text summarization. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'02)* (pp. 105–112). Tampere, Finland.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Addison Wesley.

Buckley, C., & Walz, J. (1999) The smart/empire tipster ir system. In *Proceedings of TIPSTER Phase III Workshop*.

Chuang, W. T., & Yang, J. (2000). Extracting sentence segments for text summarization: A machine learning approach. The use of unlabeled data to improve supervised learning for text summarization. In *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrival (SIGIR'00)* (pp. 152–159). New Orleans, USA.

Frankes, W. B., & Baeza-Yates, R. (1992). *Information retrieval: Data structure & algorithms*. Prentice-Hall.

Gong, Y., & Liu, X. (2001). 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 retrival (SIGIR'01)* (pp. 19–25). New Orleans, USA.

Hoa, T. D. (2005). Overview of DUC 2005. In *Proceedings of the Document Understanding Conference (DUC'05)*.

Kruengkrai, C., & Jaruskulchai, C. (2003). Generic text summarization using local and global properties of sentences. In *Proceedings of the IEEE/WIC international conference on web intelligence (IEEE/WIC'03)*.

Kupiec, J., Pedersen, J., & Chen, F. (1995). A trainable document summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'95)* (pp. 68–73). Seattle, WA, USA.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*, 788–791.

Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems, 13*, 556–562.

Li, W., Li, B., & Wu, M. (2006). Query focus guided selection strategy for DUC 2006. In *Proceedings of the Document Understanding Conference (DUC'06)*.

Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Proceedings of workshop on text summarization branches out, post-conference workshop of ACL*.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal*, 159–165.

Marcu, D. (1999). The automatic construction of large-scale corpora for summarization research. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval (SIGIR'99)* (pp. 137–144), Berkeley, CA, USA.

Mani, I., & Maybury, M. T. (1999). *Advances in automatic text*. The MIT Press.

Mani, I. (2001). *Automatic summarization*. John Benjamins Publishing Company.

Mihalcea, R. (2005). Language independent extractive summarization. In *Proceedings of the 25th National Conference on Artificial Intelligence (AAAI'05)*. Pittsburgh, PA, USA.

Nomoto, T., & Yuji, M. (2001). A new approach to unsupervised text summarization. In *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'01)* (pp. 26–34). New Oreans, LA, USA.

Shen, D., Sun, J. T., Li, H., Yang, Q., Chen, Z. (2007). Document summarization using conditional random fields. In *Proceedings of the 20th international joint conference on artificial intelligence (IJCAI'07)* (pp. 2862–2867). Hyderabad, India.

Wild, S., Curry, J., Dougherty, A. (2003). Motivating non-negative matrix factorizations. In *Proceedings of SIAM ALA*.

Xu, W., Liu, X., & Gong, Y. (2003) Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrival (SIGIR'03)* (pp. 267–273). Toronto, Canada.

Yeh, J. Y., Ke, H. R., Yang, W. P., & Meng, I. H. (2005). Text summarization using a trainable summarizer and latent semantic analysis. *Information Processing and Management, 41*, 75–95.

Zha, H. (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 retrival (SIGIR'02)* (pp. 113–120). Tampere, Finland.