

iDVS: An Interactive Multi-document Visual Summarization System

Yi Zhang, Dingding Wang, and Tao Li

School of Computer Science,
Florida International University,
Miami, FL 33199, USA
{yzhan004, dwang003, taoli}@cs.fiu.edu

Abstract. Multi-document summarization is a fundamental tool for understanding documents. Given a collection of documents, most of existing multi-document summarization methods automatically generate a static summary for all the users using unsupervised learning techniques such as sentence ranking and clustering. However, these methods almost exclude human from the summarization process. They do not allow for user interaction and do not consider users' feedback which delivers valuable information and can be used as the guidance for summarization. Another limitation is that the generated summaries are displayed in textual format without visual representation. To address the above limitations, in this paper, we develop iDVS, a visualization-enabled multi-document summarization system with users' interaction, to improve the summarization performance using users' feedback and to assist users in document understanding using visualization techniques. In particular, iDVS uses a new semi-supervised document summarization method to dynamically select sentences based on users' interaction. To this regard, iDVS tightly integrates semi-supervised learning with interactive visualization for document summarization. Comprehensive experiments on multi-document summarization using benchmark datasets demonstrate the effectiveness of iDVS, and a user study is conducted to evaluate the users' satisfaction.

Keywords: interactive multi-document summarization, visualization.

1 Introduction

Multi-document summarization aims to generate a compressed summary by extracting information from a collection of documents sharing the same or similar topics. Since the Internet faces the data overload threat with the explosive increase of the documents, automatic multi-document summarization has attracted much attention and various summarization applications have emerged in recent years.

Current research on multi-document summarization often treats it as an unsupervised learning problem. Existing summarization methods usually involve sentence clustering and ranking to extract the most important sentences from the document collection to form the summaries automatically. Although these methods can efficiently generate short summaries of the documents, they have two major limitations:

- **Lack of User Participation:** Summarization is a subjective task, i.e., different users may have different understanding and opinions on the same set of documents. However, most existing summarization methods exclude human from the summarization process, which is efficient in terms of reducing users’ workload, but it is not desired since the generated summaries is user-independent, contradicting to the subjective nature of summarization. These methods do not allow for user interaction and do not consider users’ feedback which delivers valuable information and can be used as the guidance for summarization.
- **Homogeneous Result Presentation:** Current document summarization systems present their results in textual format. As the size of the document collection increases, it becomes time-consuming, cumbersome, and labor-intensive for users to quickly locate important information or glean insights from the summary.

To address the above limitations, we develop iDVS, a visualization-enabled multi-document summarization system with users’ interaction, to improve the summarization performance using users’ feedback and to assist users in document understanding using visualization techniques. As an interactive document visual summarization system, iDVS lays out the sentence relationships in the documents visually, and iteratively selects sentences to create a summary incorporating users’ feedback. In particular, an energy-based layout algorithm is used for sentence visualization, and a new semi-supervised document summarization method is proposed to dynamically select sentences based on users’ interaction.

Note that visualization plays a vital role in iDVS. First, visualization facilitates user interaction. In iDVS, visualization techniques are used to display the sentence relationships and to help users locate critical information in a large text collection. Second, visual presentation and navigation of summary results can assist users in understanding large documents efficiently. Unlike existing work in visual text analysis, which focuses either on developing new text analytic algorithms and/or novel visualization techniques, iDVS aims to tightly integrate semi-supervised learning (for dynamic sentence selection) with interactive visualization (for user exploration) for supporting effective document summarization. We conduct extensive experiments on the benchmark datasets to demonstrate the effectiveness of iDVS on multi-document summarization and also perform user studies to evaluate the users’ satisfaction.

The rest of this paper is organized as follows. Section 2 discusses the related work on multi-document summarization, semi-supervised learning and document visualization techniques. Section 3 introduces the framework of iDVS. Methods used in iDVS including the layout algorithms, paragraph selection, and semi-supervised document summarization are described in Section 4. Section 5 illustrates an example summarization process of iDVS. Section 6 demonstrates and discusses the experimental results comprehensively. A user study is conducted in Section 7. Finally, Section 8 concludes.

2 Related Work

2.1 Multi-document Summarization

Various multi-document summarization methods have been studied recently. Our discussion here focuses on extractive methods. The most commonly used methods are

centroid based, which usually rank sentences in the document collection according to their scores calculated by a set of predefined features, such as term frequency-inverse sentence frequency (TF-ISF), sentence or term position, and number of keywords [37,29,23]. Another type of methods use sentence graph representation and select sentences based on the votes from their neighbors using ideas similar to PageRank [11]. Other methods include Latent semantic analysis (LSA) based summarization [13], Non-negative matrix factorization (NMF) based summarization [33], Conditional Random Field (CRF) based summarization [31], and hidden Markov model (HMM) based method [8]. Some query-based summarization systems are also proposed [12].

Most of the existing document summarization methods are unsupervised as described above. There are also a few studies applying supervised and semi-supervised methods to document summarization. Wong et al. [35] first extracted and combined various sentence features, and then applied supervised and semi-supervised classifiers to obtain the labels for all the sentences. Although some work considers users' opinions by analyzing user comments [19], few work has been reported on iteratively making use of users' feedback to improve the quality of the generated summaries. Most of current document summarization systems deliver results in textual format only, and there are limited attempts on document summarization to present the relationships among sentences visually.

2.2 Visual Text Analysis

The work on visual text analysis can be broadly divided into five different categories: 1) **Meta-data visualization** methods which focus on visualizing the meta data of text documents. For example, many techniques have been developed to visualize the relationships between the email senders and receivers in email corpora [28,26,21], and TileBars is used to visualize document length and query term frequency [16]. 2) **Document visualization** methods which focus on displaying document relationships. Typical systems for document visualization include the Galaxy of News [30], Jigsaw [32], and ThemeRiver [15]. 3) **Word visualization** methods which mainly show the text information at the text level. Typical systems for word visualization include TextArc (www.textarc.org), WordTree [34], and FeatureLens [10]. 4) **Text Visual analytic** methods which typically integrate visualization with some text analytic methods. For example, Ando et al. [3] developed a visualization-enabled multi-document summarization by iterative residual rescaling, Allan et al. [2] used the cluster visualization to help a user rapidly identify relevant documents, Liu et al. [25] presented an interactive, visual text analysis tool to support both top-down and bottom-up text analysis. 5) **Visual Summarization** methods which present a document summary using representative images. Google has released the "Image from the page" feature in its web search system using the images in a page to summarize the page. Jiao et al. [20] propose a visual summarization system using external images (which are not the images in the page) for search and re-finding tasks.

Our iDVS can be viewed as a text visual analytic method by integrating summarization and visualization and its focus is on generating better summary results with user interaction and feedbacks. Compared with existing text visualization systems, iDVS

tightly combines semi-supervised learning techniques with interactive visualization to support user-centric document summarization. On one hand, iDVS uses visualizations to display summarization results and facilitates user interaction. On the other hand, it employs semi-supervised learning for users to perform dynamic sentence selection.

2.3 General Visualization Analytic

In general information visualization, researchers have developed a number of methods for visual analytics. For example, HD-Eye [18] and n23Tool [36] integrate visualization and clustering algorithms by representing the possible clusters and the relationship between clusters in each projection of any interactive system to visualize and analyze clusters. A visual framework VISTA is developed in [6] to visualize multi-dimensional datasets in a 2D star-coordinate space and to validate and refine the cluster structure. Chen and Liu [7] extended VISTA to an interactive visualization-based framework for clustering large datasets. More visualization techniques can be found in [14]. In our work, we focus on visualizing text data especially summary results. In addition, iDVS provides users with visual interactions tools to dynamic select sentences from multiple perspective.

2.4 Semi-Supervised Learning

In this paper, we treat document summarization given users' guidance as a problem of semi-supervised learning. Semi-supervised learning [5] learns from a set of partially labeled data points, including both labeled and unlabeled data, to predict the labels of the unlabeled data. For the document summarization task, we ask users to read and label a small portion of the documents, thus a semi-supervised algorithm can be developed to proceed document summarization with only a small portion of prior knowledge.

3 System Framework

Figure 1 shows the framework of iDVS. The system creates short summaries by selecting sentences iteratively based on the results of the semi-supervised document summarization engine and the feedback provided by users. Each iteration of the sentence selection includes the following steps.

- (1) The system generates a 2-D view graph of current sentence set of the given documents, in which each node represents a sentence, and the location and color of the sentence are determined by the layout and clustering algorithms respectively. Here, the initial "current sentence set" contains all the sentences in the documents.
- (2) Based on the layout results, the system picks out the most important sentence and asks a user to read the paragraph where the sentence locates.
- (3) The user partially orders the sentences contained in the paragraph based on their understanding of the content in the paragraph.
- (4) A semi-supervised sentence ranking algorithm by making use of the sentence preferences from the user is then performed to rank all the sentences in the documents. Top n sentences are then selected as candidates and recommended to the user. The number n is determined by the difference between the required summary length and the length of sentences already in the summary.

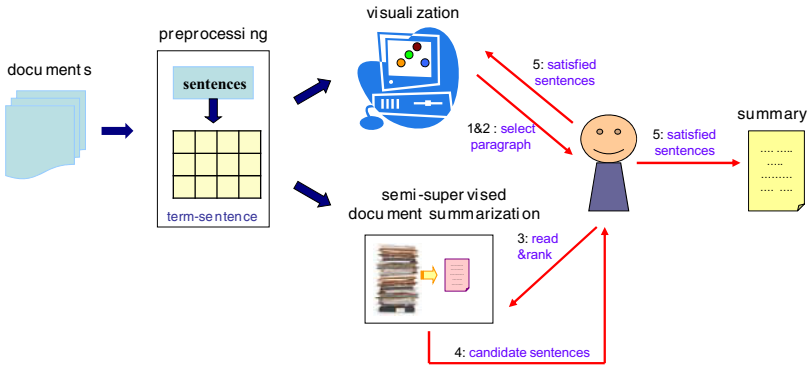


Fig. 1. The system overview

- (5) The user selects the candidate sentences which he/she is satisfied with, then these sentences are included in the summary. The selected sentences and the clusters that they belong to are then removed from current sentence set.
- (6) A new iteration starts until the required length of the summary is reached. One thing worth mentioning is that the number of sentence clusters and the number of sentences in a summary are small in general, so the involved amount of user efforts is acceptable in practice.

4 Methodology

In this section, we introduce the key techniques used in iDVS for sentence cluster layout, semi-supervised document summarization, and user interaction.

4.1 Sentence Cluster Layout

Sentence Graph Representation. Given a collection of documents, we first decompose them into sentences. An undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is then constructed to represent the relationships among the sentences, where \mathcal{V} is the vertex set and \mathcal{E} is the edge set. Each vertex in \mathcal{V} is a sentence, and each edge in \mathcal{E} is associated with the cosine similarity between two sentences (vertices). Two vertices are connected if their cosine similarity is greater than 0.

Linlog Layout Algorithm. Here, we use Linlog, a popular energy-based layout algorithm[27], to display the sentence relationships and present the clusters in the separated shapes. The energy function in Linlog is

$$E(p) = \sum_{\{u,v\}:u \neq v} (\omega_{\{u,v\}} \|p_u - p_v\| - d_v d_u \ln \|p_u - p_v\|),$$

where $\omega_{\{u,v\}}$ is the weight of the edge connecting vertices u and v , and d_u and d_v are the degrees of u and v respectively. The optimal positions p of all the vertices are obtained by minimizing \mathcal{E} .

Clustering with Maximum Modularity. The positions of nodes (sentences) displayed by the energy-based layout algorithm are consistent with the clustering results obtained by maximizing graph modularity.

Modularity [27,1] can be defined as

$$\sum_{c \in C} \left[\frac{w_c}{w_C} - \left(\frac{d(c)}{d(C)} \right)^2 \right],$$

where w_c, w_C are the sum of edge weights in cluster c and cluster set C respectively, and $d(c)$ and $d(C)$ are the sum of node degrees for all the nodes in cluster c and cluster set C .

The clustering results can be easily obtained by a bottom-up algorithm, in which each sentence is treated as a singleton cluster at the outset and then successively merge pairs of clusters until the maximum modularity is reached.

4.2 Paragraph Selection and Users' Annotation

The visualization in iDVS clearly illustrates the following information for users. (1) The radius of each node is determined by the sentence's degree. The larger the node, the more important the corresponding sentence. In other words, the largest node in a cluster corresponds to the most important sentence in that cluster. (2) Large nodes in the overlapping area of two clusters may be the transition sentences between the clusters. (3) The larger the distance between two clusters, the dissimilar the two topics. Based on the visualization of sentence relationships, the user finds the largest nodes in each cluster, then picks any one of them, and the system returns its corresponding sentence with the paragraph that the sentence belongs to for the user to read and annotate.

In the annotation, the user is asked to read the selected paragraph and provide the comparable sentence pairs in the paragraph and their preferences of the two sentences based on their understanding of the content in the paragraph. The user finally provides all the comparable sentence pairs and label their preference of the two sentences for each sentence pair in the paragraph. Here the "comparable sentence pair" means that two sentences share similar topics. Thus, we have a number of sentence pairs and their preference, e.g. (s_i, s_j) and their relationship $r_i \geq r_j$ representing sentence i is preferred to be included in the summary comparing to sentence j . We treat the sentence pair (s_i, s_j) with their preference $r_i \geq r_j$ as a "sentence preference pair".

4.3 Semi-Supervised Document Summarization

For the summarization task, we are given a set of sentences $\mathcal{S} = \{s_i\}_{i=1}^m$ associated with their importance $\{r_i\}_{i=1}^m$. Through users' interactions, we know some pairwise relationships

$$\mathcal{C} = \{(s_{k_i}, s_{k_j})\}_{k=1}^K,$$

such that if $(s_{k_i}, s_{k_j}) \in \mathcal{C}$, then $r_{k_i} \geq r_{k_j}$, where k is the index of the constraint and K is the total number of constraints. To apply the semi-supervised summarization method, we should first have a sentence graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ constructed as in Section 4.1. Then our semi-supervised model assumes that the estimated sentences preferences $\{r_i\}$ should vary sufficiently smooth with respect to \mathcal{G} . Usually such smoothness can be calculated by

$$\delta = \sum_{s_i \sim s_j} w_{ij} \left(\frac{r_i}{\sqrt{d_i}} - \frac{r_j}{\sqrt{d_j}} \right)^2. \quad (1)$$

Here $d_i = \sum_j w_{ij}$ is the degree of sentence s_i , and $s_i \sim s_j$ denotes that s_i and s_j are adjacent (i.e., there is an edge linking them). Written in its matrix form, Eq.(1) can be transformed to

$$\delta = \mathbf{r}^T \mathbf{L} \mathbf{r}, \quad (2)$$

where

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}, \quad (3)$$

is the normalized Laplacian matrix [38], and \mathbf{I} is an $m \times m$ identity matrix. Some theoretical research [4][17] showed that if we assume the sentences in \mathcal{S} are sampled from some continuous ‘‘sentence manifold’’ $\mathcal{M}_{\mathcal{S}}$, then when the number of sampled sentences tends to infinity, the normalized Laplacian will converge to the Laplace-Beltrami operator on $\mathcal{M}_{\mathcal{S}}$, which is strongly related to the functional smoothness defined on $\mathcal{M}_{\mathcal{S}}$.

In our semi-supervised summarization scenario, we assume that we have an ‘‘initial’’ preference vector \mathbf{r}_0 which is estimated via some unsupervised methods, and our method can be viewed as a way to ‘‘improve’’ such initial guess \mathbf{r}_0 by (1) the sentence-sentence geometric structure; (2) prior knowledge on sentence preferences. For simplicity, we use Euclidean distance to measure the discordance of the final preference vector \mathbf{r} and initial \mathbf{r}_0 whose objective is to solve the following optimization problem

$$\min_{\mathbf{r}} \|\mathbf{r} - \mathbf{r}_0\|^2 + \lambda \mathbf{r}^T \mathbf{L} \mathbf{r}, r_{k_i} \geq r_{k_j}, \forall (s_{k_i}, s_{k_j}) \in \mathcal{C}, \quad (4)$$

where $\mathbf{r} = [r_1, r_2, \dots, r_m]^T$ and has been normalized to 1, and $\lambda > 0$ is the regularization parameter to tradeoff the loss and smoothness terms. If we denote $\mathbf{e}_{k_i} \in \mathbf{R}^{m \times 1}$ as an all-zero vector with only its k_i -th element being 1, then we can rewrite problem (4) in its standard form as

$$\min_{\mathbf{r}} \|\mathbf{r} - \mathbf{r}_0\|^2 + \lambda \mathbf{r}^T \mathbf{L} \mathbf{r}, \mathbf{r}^T (\mathbf{e}_{k_i} - \mathbf{e}_{k_j}) \geq 0, \forall (s_{k_i}, s_{k_j}) \in \mathcal{C}. \quad (5)$$

Clearly, the objective of the above problem is quadratic in \mathbf{r} and the constraints are linear in \mathbf{r} , which makes it a convex quadratic programming problem whose global optimum can be easily found by some mature numerical methods.

The objective of problem (5) can also be understood from a probabilistic perspective, where we have the initial guess \mathbf{r}_0 , the sentences \mathcal{S} , and we want to learn the true sentence preferences \mathbf{r} . In general, the optimal \mathbf{r} can be obtained as the one with the maximum posterior probability

$$P(\mathbf{r} | \mathcal{S}, \mathbf{r}_0) = \frac{1}{Z} P(\mathbf{r} | \mathcal{S}) P(\mathbf{r} | \mathbf{r}_0),$$

where

$$P(\mathbf{r}|\mathcal{S}) = \frac{1}{Z_1} \exp(-\mathbf{r}^T \mathbf{L}\mathbf{r}),$$

is the prior probability of \mathbf{r} given the sentence set \mathcal{S} ,

$$P(\mathbf{r}|\mathbf{r}_0) = \frac{1}{Z_2} \exp(-\alpha \|\mathbf{r} - \mathbf{r}_0\|^2)$$

is the likelihood of \mathbf{r} given the initial guess \mathbf{r}_0 , and Z_1, Z_2, Z, α are constants. If we define the energy E as

$$E = \mathbf{r}^T \mathbf{L}\mathbf{r} + \alpha \|\mathbf{r} - \mathbf{r}_0\|^2.$$

Then maximizing the posterior $P(\mathbf{r}|\mathcal{X}, \mathbf{r}_0)$ is equivalent to minimizing the energy E . Let $\lambda = \frac{1}{\alpha}$, we can get the objective in problem (5). Then we sort \mathbf{r} in descending order to get the ranking of the sentences in the document collection, and the top sentences are selected as the candidate sentences and returned to the user.

5 An Illustrative Example

Figure 2 illustrates an example of the visualization results in each iteration of the summary generation process. Some key features of iDVS can be demonstrated in this example as follows.

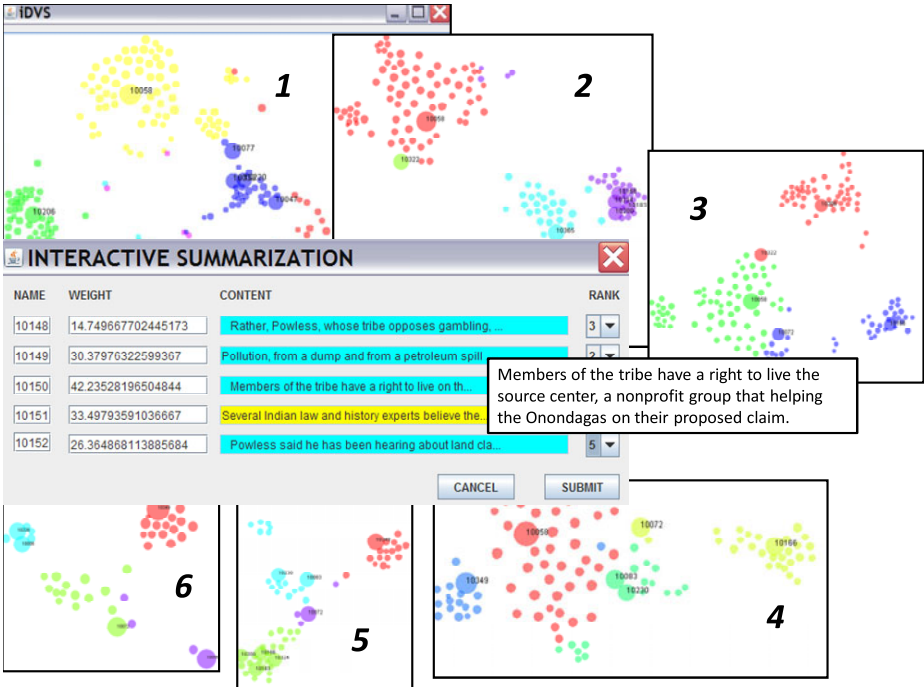


Fig. 2. An illustrative example

- Given a set of 25 news articles, the system demonstrates the sentence clusters using the layout and clustering algorithms described in Section 4.1. The layout results of the original sentence set are shown in the upper left part of Figure 2. The most important sentences (nodes) in each cluster are of large size and labeled with their sentence IDs. In this example, there exist six clusters in the original sentence set, each of which is associated with a unique color.
- As the sizes of nodes and clusters indicate the importance of a sentence and a sentence cluster, users select one node from the graph, the system returns the corresponding sentence and the paragraph it belongs to, each sentence in the graph is displayed in a panel as shown in the lower right part in Figure 2. Note that sentences in the same paragraph are not necessary to be in the same cluster, thus different colors in the panel indicate the sentence clusters. For the space limit, only partial of the sentences are displayed, and if users are interested in any sentences, they can move the mouse over the sentence, and a tip will pop up to show the entire sentence. Users can read and rank the sentences and then submit their preferences through the system to the semi-supervised document summarization engine.
- In each iteration, the summarization engine will generate candidate sentences using the proposed semi-supervised document summarization method (as described in Section 4.3). Users can pick their interested sentences to include in the summary.
- Then the system will remove those sentences and their clusters which have been added into the summary, and a new iteration starts. Note that if the user is not satisfied with any candidate sentence, no nodes will be removed and the layout won't change. However, the user can either re-rank the sentences in the most important paragraph or the system can select the second important paragraph for users to annotate. In this example, there are three iterations, and from the figure, we clearly observe the reduction of number of sentences in each iteration.

6 Experiments

6.1 Data Description

To evaluate the summarization results empirically, we use the DUC2002 and DUC2004 data sets, both of which are open benchmark data sets from Document Understanding Conference (DUC) for generic automatic summarization evaluation. Table 1 gives a brief description of the data sets.

Table 1. Description of the data sets for multi-document summarization

	DUC2002	DUC2004
number of document collections	59	50
number of documents in each collection	~10	10
data source	TREC	TDT
summary length	200 words	665bytes

6.2 Implemented Summarization Systems

We implement the following unsupervised and supervised document summarization methods as the baseline systems to examine the quality of the summaries generated by our proposed method: (1) **Random**: selects sentences randomly for each document collection. (2) **Centroid**: applies MEAD algorithm proposed in [29] using centroid value, positional value, and first-sentence overlap as features. (3) **LexPageRank**: a graph-based summarization method recommending sentences by the voting of their neighbors [11]. (4) **LSA**: conducts latent semantic analysis on terms by sentences matrix as proposed in [13]. (5) **NMF**: performs NMF on terms by sentences matrix and ranks the sentences by their weighted scores [22]. (6) **KM**: performs K-means algorithm to clustering the sentences and chooses the center sentences in each clusters. (7) **Supervised**: transfers the document summarization problem to a two-class sentence classification problem (i.e. one class represents the sentences included in the summary, while the other class represents the rest of the sentences). Then Naive Bayesian classifier (NBC) is used as the learning approach. The class label information comes from one human summary in DUC for evaluation [35]. (8) **iDVS**: our proposed semi-supervised document summarization method as described in Section 4.

Table 2. Overall performance comparison on DUC2002 data using ROUGE evaluation methods

Systems	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-W	ROUGE-SU
DUC Best	0.49869	0.25229	0.46803	0.20071	0.28406
Random	0.38475	0.11692	0.37218	0.15941	0.18057
Centroid	0.45379	0.19181	0.43237	0.17971	0.23629
LexPageRank	0.47963	0.22949	0.44332	0.18978	0.26198
LSA	0.43078	0.15022	0.40507	0.15220	0.20226
NMF	0.44587	0.16280	0.41513	0.16072	0.21687
KM	0.43156	0.15135	0.40376	0.15038	0.20144
Consistency Method	0.46449	0.22173	0.44514	0.18311	0.25670
Variant Consistency 1	0.49990	0.26461	0.47233	0.20118	0.28745
Variant Consistency 2	0.47121	0.23322	0.42578	0.17266	0.23007
Homonic-CMN	0.46809	0.23039	0.41759	0.17556	0.22876
Green's Function	0.46591	0.22788	0.41113	0.16092	0.22227
Supervised	0.51243	0.26538	0.47731	0.20827	0.29634
iDVS	0.50612	0.25981	0.47535	0.20279	0.28863

Five Semi-supervised Summarization Systems: In addition to the above systems, we also compare our *interactive* semi-supervised document summarization system (iDVS) with the following baseline summarization systems which directly make use of semi-supervised learning methods. In these baseline systems, the most important sentence from each sentence cluster obtained using clustering with maximum modularity (as described in Section 4.2) is labeled as summary sentence. Then semi-supervised learning approaches are used to generate a cluster C with size n_C whose elements are the candidate summary sentences. Finally, a sentence subset K with size n_K is selected from C to form the final summary using Eq.(6):

$$\min \frac{\gamma}{n_C \times n_K} \sum_{s_i \in K} [\sum_{s_j \in C} Sim(s_i, s_j)] - \frac{1 - \gamma}{n_K \times n_K} \sum_{s_i, s_j \in K} Sim(s_i, s_j), \quad (6)$$

where the first term is the average similarity between the selected sentence and the rest sentences and the second term is the average pairwise sentence similarity in K , both are calculated using cosine similarity. γ is a weight parameter, and is set to 0.6 empirically. In other words, the final sentence selection aims to choose sentences that have high similarity values with all other sentences in C and low similarity values with the selected sentences. Five popular semi-supervised learning approaches are used in the baseline systems: (1-3) the algorithm proposed in [35] which conducts semi-supervised learning with local and global consistency (**Consistency Method**), and two of its variants (**Variants Consistency 1** and **Variants Consistency 2**); (4) Harmonic Gaussian field method coupled with the Class Mass Normalization (**Harmonic-CMN**) [39]; (5) Green's function learning algorithm (**Green's Function**) Proposed in [9].

6.3 Evaluation Method

We use ROUGE [24] toolkit (version 1.5.5) to measure the summarization performance, which is widely applied by DUC for performance evaluation. It measures the quality of a summary by counting the unit overlaps between the candidate summary and a set of reference summaries. Intuitively, the higher the scores, the more similar between the two summaries. As we have similar conclusions for different scores, for simplicity, in this paper, we only report the average F-measure scores generated by ROUGE-1, ROUGE-2, ROUGE-L, ROUGE-W and ROUGE-SU to compare the implemented systems. One thing worth mentioning is that since different users might have different perceptions of which sentence is more important, the results of iDVS reported in the experiments are the average scores obtained from three users.

6.4 Experimental Results

Overall Performance. First of all, we compare the overall performance of our proposed semi-supervised document summarization method with other most widely used unsupervised and supervised summarization methods. Table 2 and Table 3 show the ROUGE evaluation results on DUC2002 and DUC2004 data sets respectively. We set λ to be 0.6 empirically. From the results, we have the following observations: (1) Our iDVS method outperforms all the unsupervised summarization methods. This benefits from the guidance of the preference and feedback provided by users and our proposed semi-supervised learning method which makes use of the user interaction. (2) We compare our method with the best team from DUC competition. Although the advanced natural language processing techniques are used by the best team, our semi-supervised method still outperforms it, which demonstrates the effectiveness of our method utilizing a small amount of labeled data. (3) Compared with the baseline summarization systems which directly make use of semi-supervised learning, our iDVS method is able to utilize user interactions and achieves the best performance. The performance

improvement clearly demonstrate the advantage of user participation in our system. (4) Our method maintains the comparable performance with the supervised summarization approach which makes use of the human generated summaries. Although the results from supervised learning method are encouraging, a fully labeled training data is too expensive and usually infeasible to obtain.

To better demonstrate the results, Figure 3 visually illustrate the comparison. As we have similar conclusion on different ROUGE scores, due to the space limit, we only show the ROUGE-1 results in these figures.

Table 3. Overall performance comparison on DUC2004 data using ROUGE evaluation methods

Systems	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-W	ROUGE-SU
DUC Best	0.38224	0.09216	0.38687	0.13325	0.13233
Random	0.31865	0.06377	0.34521	0.11734	0.11779
Centroid	0.36728	0.07379	0.36182	0.12439	0.12511
LexPageRank	0.37842	0.08572	0.37531	0.13121	0.13097
LSA	0.34145	0.06538	0.34973	0.12042	0.11946
NMF	0.36747	0.07261	0.36749	0.12961	0.12918
KM	0.34872	0.06937	0.35882	0.12339	0.12115
Consistency Method	0.37565	0.09776	0.37781	0.12765	0.13000
Variant Consistency 1	0.39225	0.10477	0.39547	0.13762	0.13651
Variant Consistency 2	0.37740	0.09803	0.38006	0.12812	0.13091
Homonic-CMN	0.37211	0.09466	0.37094	0.12233	0.12997
Green's Function	0.36411	0.07215	0.36645	0.12470	0.12261
Supervised	0.39622	0.10973	0.39956	0.14221	0.14095
iDVS	0.39227	0.10524	0.39633	0.13816	0.13751

Paragraph Selection. In the experiments, we select paragraphs based on users' feedback and the visualized structure of sentences. In this set of experiments, we examine this paragraph selection method and compare it with the following selection solutions: (a) **RandomSel**: randomly selects paragraphs of the documents; and (b) **IdealSel**: selects paragraphs containing the sentences in the human summaries used for evaluation. Figure 4 demonstrates the comparison results. From the results, we can see that our paragraph selection method is more effective than the random selection. Although the ideal selection performs the best, in real application the human summaries are infeasible to obtain, while our selection method is simple and cost efficient.

Parameter Tuning. In Figure 5, we gradually tune the parameter λ in our method to adjust the weights between the two parts of the objective function in 5. We vary λ from 0.1 to 0.9 in every 0.1 interval. From the results, we can see that when λ is 0.6, the generated summaries achieve the highest performance.

7 A User Survey

To better evaluate the results of iDVS, we conduct a user survey. The subjects of the survey are fifteen students at different levels and from various majors of a university.

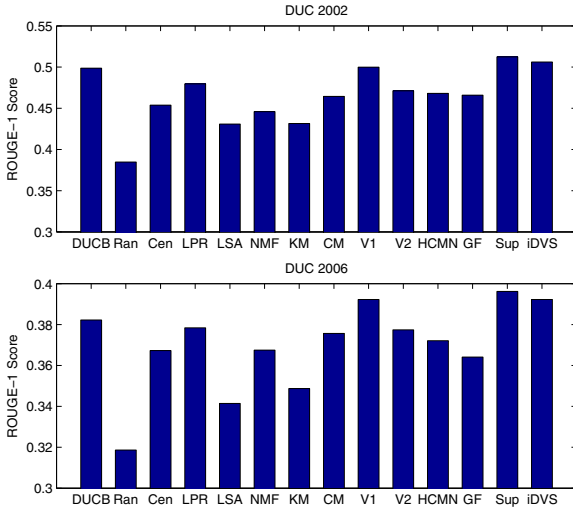


Fig. 3. Overall summarization performance on DUC2002 data (Top), DUC2006 data (Bottom) using ROUGE-1. DUCB, Ran, Cen, LPR, LSA, NMF, CM, V1, V2, HCMN, GF, Sup are corresponding to the summarization methods named DUC Best, Random, Centroid, LexPageRank, LSA, NMF, Consistency Method, Variant Consistency 1, Variant Consistency 2, Homonic-CMN, Green’s Function and Supervised respectively.

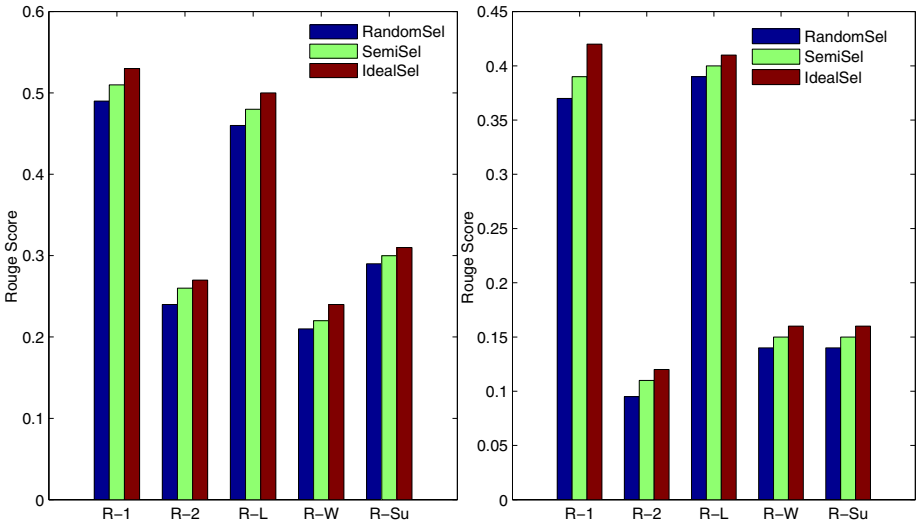


Fig. 4. iDVS results with different paragraph selection methods on DUC2002 data (left) and DUC2004 data (right)

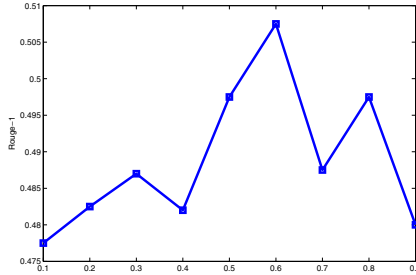


Fig. 5. iDVS parameter tuning using DUC2002 data

Table 4. User satisfaction comparison

Systems	Scores
iDVS	4.07
Random	1.47
Centroid	2.53
LexPageRank	2.80
Supervised	3.47

Each participant randomly selects a set of news documents, and use iDVS to form a summary. Then they are asked to assign a score of 1 (least satisfaction) to 5 (highest satisfaction), according to their satisfaction of the use of iDVS. We compare iDVS with some typical summarization systems including Random, Centroid, LexPageRank, and Supervised. The first three methods are unsupervised methods, and the last one is supervised method. And all of these systems do not support visualization of the sentence relationships. Table 4 demonstrate the satisfaction scores for each system, and the results show the effectiveness and high usability of iDVS. From the results, we have the following observations: (1) Generally, users' satisfaction scores are consistent with the performance of the summarization systems. (2) Users are highly satisfied with iDVS because (1) the generated summaries are based on the guidance and feedback through users' interaction, i.e., it supports personalized opinions; (2) iDVS provides visualization to illustrate the latent relationships among sentences.

8 Conclusion

In this paper, we develop iDVS, an interactive document visual summarization system, to visualize the structure of sentences contained in a document collection, and generate a short summary based on users' opinions iteratively. The main contributions of iDVS are that (1) the visual system can help users better understand the latent relationships in the documents/sentences easily; (2) the summarization process incorporates users' feedback including some domain knowledge and personalized opinions, which can validate and refine the summarization results effectively. Comprehensive experiments and a user study demonstrate the effectiveness of iDVS.

Acknowledgement. The work is partially supported by US National Science Foundation under grants IIS-0546280 and CCF-0830659.

References

1. Agarwal, G., Kempe, D.: Modularity-maximizing graph communities via mathematical programming. *The European Physical Journal B - Condensed Matter and Complex Systems* 66(3), 409–418 (2008)
2. Allan, J., Leouski, A.V., Swan, R.C.: Interactive cluster visualization for information retrieval. In: *ECDL* (1998)
3. Ando, R., Boguraev, B., Byrd, R., Neff, M.: Visualization-enabled multi-document summarization by iterative residual rescaling. *Nat. Lang. Eng.* 11(1), 67–86 (2005)
4. Belkin, M., Niyogi, P.: Towards a theoretical foundation for laplacian-based manifold methods. In: Auer, P., Meir, R. (eds.) *COLT 2005. LNCS (LNAI)*, vol. 3559, pp. 486–500. Springer, Heidelberg (2005)
5. Chapelle, O., Schölkopf, B., Zien, A. (eds.): *Semi-Supervised Learning*. MIT Press, Cambridge (2006)
6. Chen, K., Liu, L.: Vista: validating and refining clusters via visualization. *Information Visualization* 3(4), 257–270 (2004)
7. Chen, K., Liu, L.: ivibrate: Interactive visualization-based framework for clustering large datasets. *ACM Trans. Inf. Syst.* 24(2), 245–294 (2006)
8. Conroy, J., O’Leary, D.: Text summarization via hidden markov models. In: *SIGIR*, pp. 406–407 (2001)
9. Ding, C., Jin, R., Li, T., Simon, H.D.: A learning framework using green’s function and kernel regularization with application to recommender system. In: *SIGKDD* (2007)
10. Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B., Plaisant, C.: Discovering interesting usage patterns in text collections: integrating text mining with visualization. In: *CIKM*, pp. 213–222 (2007)
11. Erkan, G., Radev, D.: Lexpagerank: Prestige in multi-document text summarization. In: *EMNLP* (2004)
12. Goldstein, J., Kantrowitz, M., Mittal, V., Carbonell, J.: Summarizing text documents: Sentence selection and evaluation metrics. In: *SIGIR*, pp. 121–128 (1999)
13. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: *SIGIR*, pp. 75–95 (2001)
14. Grinstain, G., Ankerst, M., Keim, D.: Visual data mining: Background, applications, ad drug discovery applications. In: *SIGMOD* (1999)
15. Havre, S., Hetzler, E., Whitney, P., Nowell, L.: Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics* 8(1), 9–20 (2002)
16. Hearst, M.A.: Tilebars: visualization of term distribution information in full text information access. In: *CHI*, pp. 59–66 (1995)
17. Hein, M., Audibert, J., Von Luxburg, U.: From graphs to manifolds - weak and strong pointwise consistency of graph laplacians. In: Auer, P., Meir, R. (eds.) *COLT 2005. LNCS (LNAI)*, vol. 3559, pp. 470–485. Springer, Heidelberg (2005)
18. Hinneburg, A., Keim, D., Wawryniuk, M.: Visual mining of high-dimensional data. *IEEE Computer Graphics and Applications* (1999)
19. Hu, M., Sun, A., Lim, E.-P.: Comments-oriented document summarization: understanding documents with readers’ feedback. In: *SIGIR*, pp. 291–298 (2008)

20. Jiao, B., Yang, L., Xu, J., Wu, F.: Visual summarization of web pages. In: SIGIR, pp. 499–506 (2010)
21. Kerr, B.: Thread arcs: an email thread visualization. In: InfoVis, pp. 211–218 (2003)
22. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS (2001)
23. Lin, C.-Y., Hovy, E.: From single to multi-document summarization: A prototype system and its evaluation. In: ACL, pp. 457–464 (2001)
24. Lin, C.-Y., Hovy, E.: Automatic evaluation of summaries using n-gram co-occurrence statistics. In: NLT-NAACL, pp. 71–78 (2003)
25. Liu, S., Zhou, M.X., Pan, S., Qian, W., Cai, W., Lian, X.: Interactive, topic-based visual text summarization and analysis. In: CIKM, pp. 543–552 (2009)
26. Nardi, B.A., Whittaker, S., Isaacs, E., Creech, M., Johnson, J., Hainsworth, J.: Integrating communication and information through contactmap. *Commun. ACM* 45(4), 89–95 (2002)
27. Noack, A.: Modularity clustering is force-directed layout. *Physical Review E* 79, 026102 (2009)
28. Perer, A., Smith, M.A.: Contrasting portraits of email practices: visual approaches to reflection and analysis. In: AVI 2006, pp. 389–395 (2006)
29. Radev, D., Jing, H., Stys, M., Tam, D.: Centroid-based summarization of multiple documents. In: Information Processing and Management, pp. 919–938 (2004)
30. Rennison, E.: Galaxy of news: an approach to visualizing and understanding expansive news landscapes. In: UIST 1994, pp. 3–12 (1994)
31. Shen, D., Sun, J.-T., Li, H., Yang, Q., Chen, Z.: Document summarization using conditional random fields. In: IJCAI, pp. 2862–2867 (2007)
32. Stasko, J., Görg, C., Liu, Z.: Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization* 7(2), 118–132 (2008)
33. Wang, D., Li, T., Zhu, S., Ding, C.H.Q.: Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In: SIGIR, pp. 307–314 (2008)
34. Wattenberg, M., Viégas, F.B.: The word tree, an interactive visual concordance. *IEEE Transactions on Visualization and Computer Graphics* 14(6), 1221–1228 (2008)
35. Wong, K.-F., Wu, M., Li, W.: Extractive summarization using supervised and semi-supervised learning. In: Coling (2008)
36. Yang, L.: n23tool: A tool for exploring large relational datasets through 3d dynamic projections. In: CIKM (2000)
37. Yih, W.-T., Goodman, J., Vanderwende, L., Suzuki, H.: Multi-document summarization by maximizing informative content-words. In: IJCAI, pp. 1776–1782 (2007)
38. Zhou, D., Bousquet, O., Navin Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: NIPS, vol. 16, pp. 321–328 (2004)
39. Zhu, X.: Semi-supervised learning literature survey. Technical report, Computer Sciences, University of Wisconsin-Madison (2005)