Multi-Document Summarization via the Minimum Dominating Set

Chao Shen and **Tao Li**
School of Computing and Information Sciences
Florida Internation University
{cshen001|taoli}@cs.fiu.edu

Abstract

Multi-document summarization has been an important problem in information retrieval. It aims to distill the most important information from a set of documents to generate a compressed summary. Given a sentence graph generated from a set of documents where vertices represent sentences and edges indicate that the corresponding vertices are similar, the extracted summary can be described using the idea of graph domination. In this paper, we propose a new principled and versatile framework for multi-document summarization using the minimum dominating set. We show that four well-known summarization tasks including generic, query-focused, update, and comparative summarization can be modeled as different variations derived from the proposed framework. Approximation algorithms for performing summarization are also proposed and empirical experiments are conducted to demonstrate the effectiveness of our proposed framework.

1 Introduction

As a fundamental and effective tool for document understanding and organization, multi-document summarization enables better information services by creating concise and informative reports for a large collection of documents. Specifically, in multi-document summarization, given a set of documents as input, the goal is to produce a condensation (i.e., a generated summary) of the content of the entire input set (Jurafsky and Martin, 2008). The generated summary can be generic where it simply gives the important information contained in the input documents without any particular information needs or query/topic-focused where it is produced in response to a user query or related to a topic or concern the development of an event (Jurafsky and Martin, 2008; Mani, 2001).

Recently, new summarization tasks such as update summarization (Dang and Owczarzak, 2008) and comparative summarization (Wang et al., 2009a) have also been proposed. Update summarization aims to generate short summaries of recent documents to capture new information different from earlier documents and comparative summarization aims to summarize the differences between comparable document groups.

In this paper, we propose a new principled and versatile framework for multi-document summarization using the *minimum dominating set*. Many known summarization tasks including generic, query-focused, update, and comparative summarization can be modeled as different variations derived from the proposed framework. The framework provides an elegant basis to establish the connections between various summarization tasks while highlighting their differences.

In our framework, a sentence graph is first generated from the input documents where vertices represent sentences and edges indicate that the corresponding vertices are similar. A natural method for describing the extracted summary is based on the idea of graph domination (Wu and Li, 2001). *A dominating set* of a graph is a subset of vertices such that every vertex in the graph is either in the subset or adjacent to a vertex in the subset; and

Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 984–992,
Beijing, August 2010

a minimum dominating set is a dominating set with the minimum size. The minimum dominating set of the sentence graph can be naturally used to describe the summary: it is *representative* since each sentence is either in the minimum dominating set or connected to one sentence in the set; and it is with *minimal redundancy* since the set is of minimum size. Approximation algorithms are proposed for performing summarization and empirical experiments are conducted to demonstrate the effectiveness of our proposed framework. Though the dominating set problem has been widely used in wireless networks, this paper is the first work on using it for modeling sentence extraction in document summarization.

The rest of the paper is organized as follows. In Section 2, we review the related work about multi-document summarization and the dominating set. After introducing the minimum dominating set problem in graph theory in Section 3, we propose the minimum dominating set based framework for multi-document summarization and model the four summarization tasks including generic, query-focused, update, and comparative summarization in Section 4. Section 5 presents the experimental results and analysis, and finally Section 6 concludes the paper.

2 Related Work

Generic Summarization For generic summarization, a saliency score is usually assigned to each sentence and then the sentences are ranked according to the saliency score. The scores are usually computed based on a combination of statistical and linguistic features. MEAD (Radev et al., 2004) is an implementation of the centroid-based method where the sentence scores are computed based on sentence-level and inter-sentence features. SumBasic (Nenkova and Vanderwende, 2005) shows that the frequency of content words alone can also lead good summarization results. Graph-based methods (Erkan and Radev, 2004; Wan et al., 2007b) have also been proposed to rank sentences or passages based on the PageRank algorithm or its variants.

Query-Focused Summarization In query-focused summarization, the information of the given topic or query should be incorporated into summarizers, and sentences suiting the user's declared information need should be extracted. Many methods for generic summarization can be extended to incorporate the query information (Saggion et al., 2003; Wei et al., 2008). Wan et al. (Wan et al., 2007a) make full use of both the relationships among all the sentences in the documents and relationship between the given query and the sentences by manifold ranking. Probability models have also been proposed with different assumptions on the generation process of the documents and the queries (Daumé III and Marcu, 2006; Haghighi and Vanderwende, 2009; Tang et al., 2009).

Update Summarization and Comparative Summarization Update summarization was introduced in Document Understanding Conference (DUC) 2007 (Dang, 2007) and was a main task of the summarization track in Text Analysis Conference (TAC) 2008 (Dang and Owczarzak, 2008). It is required to summarize a set of documents under the assumption that the reader has already read and summarized the first set of documents as the main summary. To produce the update summary, some strategies are required to avoid redundant information which has already been covered by the main summary. One of the most frequently used methods for removing redundancy is Maximal Marginal Relevance(MMR) (Goldstein et al., 2000). Comparative document summarization is proposed by Wang et. al. (Wang et al., 2009a) to summarize the differences between comparable document groups. A sentence selection approach is proposed in (Wang et al., 2009a) to accurately discriminate the documents in different groups modeled by the conditional entropy.

The Dominating Set Many approximation algorithms have been developed for finding minimum dominating set for a given graph (Guha and Khuller, 1998; Thai et al., 2007). Kann (Kann, 1992) shows that the minimum dominating set problem is equivalent to set cover problem, which is a well-known NP-hard problem. Dominating set has been widely used for clustering in wireless networks (Chen and Liestman, 2002; Han and Jia, 2007). It has been used to find topic words for hierarchical summarization (Lawrie et al., 2001), where a set of topic words is extracted as a dominating set of word graph. In our work, we use the minimum dominating set to formalize the sentence extraction for document summarization.

3 The Minimum Dominating Set Problem

Given a graph $G =< V, E >$, *a dominating set* of G is a subset S of vertices with the following property: each vertex of G is either in the dominating set S, or is adjacent to some vertices in S.

Problem 3.1. *Given a graph G, the minimum dominating set problem (MDS) is to find a minimum size subset S of vertices, such that S forms a dominating set.*

MDS is closely related to the set cover problem (SC), a well-known NP-hard problem.

Problem 3.2. *Given F, a finite collection $\{S_1, S_2, \ldots, S_n\}$ of finite sets, the set cover problem (SC) is to find the optimal solution*

$$F^* = \arg\min_{F' \subseteq F} |F'| \ s.t. \bigcup_{S' \in F'} S' = \bigcup_{S \in F} S.$$

Theorem 3.3. *There exists a pair of polynomial time reduction between MDS and SC.*

So, MDS is also NP-hard and it has been shown that there are no approximate solutions within $c \log |V|$, for some $c > 0$ (Feige, 1998; Raz and Safra, 1997).

3.1 An Approximation Algorithm

A greedy approximation algorithm for the SC problem is described in (Johnson, 1973). Basically, at each stage, the greedy algorithm chooses the set which contains the largest number of uncovered elements.

Based on Theorem 3.3, we can obtain a greedy approximation algorithm for MDS. Starting from an empty set, if the current subset of vertices is not the dominating set, a new vertex which has the most number of the adjacent vertices that are not adjacent to any vertex in the current set will be added.

Proposition 3.4. *The greedy algorithm approximates SC within $1 + \ln s$ where s is the size of the largest set.*

It was shown in (Johnson, 1973) that the approximation factor for the greedy algorithm is no more than $H(s)$, the s-th harmonic number:

$$H(s) = \sum_{k=1}^{s} \frac{1}{k} \leq \ln s + 1$$

Corollary 3.5. *MDS has a approximation algorithm within $1 + \ln \Delta$ where Δ is the maximum degree of the graph.*

Corollary 3.5 follows directly from Theorem 3.3 and Proposition 3.4.

4 The Summarization Framework

4.1 Sentence Graph Generation

To perform multi-document summarization via minimum dominating set, we need to first construct a sentence graph in which each node is a sentence in the document collection. In our work, we represent the sentences as vectors based on tf-isf, and then obtain the cosine similarity for each pair of sentences. If the similarity between a pair of sentences s_i and s_j is above a given threshold λ, then there is an edge between s_i and s_j.

For generic summarization, we use all sentences for building the sentence graph. For query-focused summarization, we only use the sentences containing at least one term in the query. In addition, when a query q is involved, we assign each node s_i a weight, $w(s_i) = d(s_i, q) = 1 - \cos(s_i, q)$, to indicate the distance between the sentence and the query q.

After building the sentence graph, we can formulate the summarization problem using

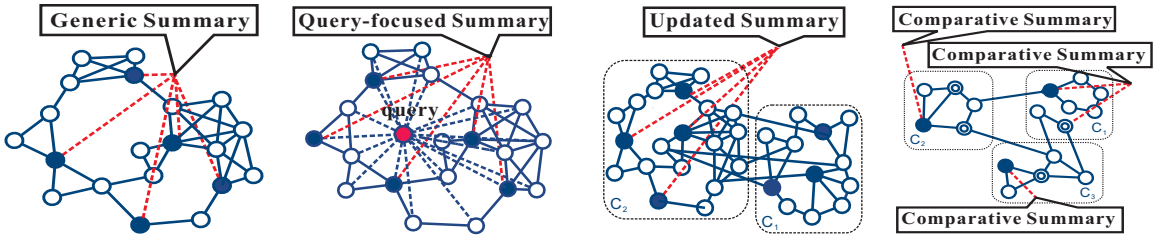

Figure 1: Graphical illustrations of multi-document summarization via the minimum dominating set. (a): The minimum dominating set is extracted as the generic summary. (b):The minimum weighted dominating set is extracted as the query-based summary. (c):Vertices in the right rectangle represent the first document set C_1, and ones in the left represent the second document set where update summary is generated. (d):Each rectangle represents a group of documents. The vertices with rings are the dominating set for each group, while the solid vertices are the complementary dominating set, which is extracted as comparative summaries.

the minimum dominating set. A graphical illustration of the proposed framework is shown in Figure 1.

4.2 Generic Summarization

Generic summarization is to extract the most representative sentences to capture the important content of the input documents. Without taking into account the length limitation of the summary, we can assume that the summary should represent all the sentences in the document set (i.e., every sentence in the document set should either be extracted or be similar with one extracted sentence). Meanwhile, a summary should also be as short as possible. Such summary of the input documents under the assumption is exactly the minimum dominating set of the sentence graph we constructed from the input documents in Section 4.1. Therefore the summarization problem can be formulated as the minimum dominating set problem.

However, usually there is a length restriction for generating the summary. Moreover, the MDS is NP-hard as shown in Section 3. Therefore, it is straightforward to use a greedy approximation algorithm to construct a subset of the dominating set as the final summary. In the greedy approach, at each stage, a sentence which is optimal according to the local criteria will be extracted. Algorithm 1 describes

Algorithm 1 Algorithm for Generic Summarization

INPUT: G, W
OUTPUT: S

$S = \emptyset$
$T = \emptyset$
while $L(S) < W$ and $V(G)! = S$ **do**
  for $v \in V(G) - S$ **do**
    $s(v) = |\{ADJ(v) - T\}|$
  $v^* = \arg\max_v s(v)$
  $S = S \cup \{v^*\}$
  $T = T \cup ADJ(v^*)$

an approximation algorithm for generic summarization. In Algorithm 1, G is the sentence graph, $L(S)$ is the length of the summary, W is the maximal length of the summary, and $ADJ(v) = \{v' | (v', v) \in E(G)\}$ is the set of vertices which are adjacent to the vertex v. A graphical illustration of generic summarization using the minimum dominating set is shown in Figure 1(a).

4.3 Query-Focused Summarization

Letting G be the sentence graph constructed in Section 4.1 and q be the query, the query-focused summarization can be modeled as

$$D^* = \arg\min_{D \subseteq G} \sum_{s \in D} d(s, q) \quad (1)$$
$$\text{s.t. } D \text{ is a dominating set of } G.$$

Note that $d(s, q)$ can be viewed as the weight of vertex in G. Here the summary length is minimized implicitly, since if $D' \subseteq D$, then

$\sum_{s\in D'} d(s,q) \leq \sum_{s\in D} d(s,q)$. The problem in Eq.(1) is exactly a variant of the minimum dominating set problem, i.e., the minimum weighted dominating set problem (MWDS).

Similar to MDS, MWDS can be reduced from the weighted version of the SC problem. In the weighted version of SC, each set has a weight and the sum of weights of selected sets needs to be minimized. To generate an approximate solution for the weighted SC problem, instead of choosing a set i maximizing $|SET(i)|$, a set i minimizing $\frac{w(i)}{|SET(i)|}$ is chosen, where $SET(i)$ is composed of uncovered elements in set i, and $w(i)$ is the weight of set i. The approximate solution has the same approximation ratio as that for MDS, as stated by the following theorem (Chvatal, 1979).

Theorem 4.1. *An approximate weighted dominating set can be generated with a size at most* $1+\log\Delta\cdot|OPT|$*, where* Δ *is the maximal degree of the graph and* OPT *is the optimal weighted dominating set.*

Accordingly, from generic summarization to query-focused summarization, we just need to modify line 6 in Algorithm 1 to

$$v^* = \arg\min_v \frac{w(v)}{s(v)}, \tag{2}$$

where $w(v)$ is the weight of vertex v. A graphical illustration of query-focused summarization using the minimum dominating set is shown in Figure 1(b).

4.4 Update Summarization

Give a query q and two sets of documents C_1 and C_2, update summarization is to generate a summary of C_2 based on q, given C_1. Firstly, summary of C_1, referred as D_1 can be generated. Then, to generate the update summary of C_2, referred as D_2, we assume D_1 and D_2 should represent all query related sentences in C_2, and length of D_2 should be minimized.

Let G_1 be the sentence graph for C_1. First we use the method described in Section 4.3 to extract sentences from G_1 to form D_1. Then we expand G_1 to the whole graph G using the second set of documents C_2. G is then the graph presentation of the document set including C_1 and C_2. We can model the update summary of C_2 as

$$D^* = \arg\min_{D_2} \sum_{s\in D_2} w(s) \tag{3}$$
$$\text{s.t. } D_2 \cup D_1 \text{ is a dominating set of } G.$$

Intuitively, we extract the smallest set of sentences that are closely related to the query from C_2 to complete the partial dominating set of G generated from D_1. A graphical illustration of update summarization using the minimum dominating set is shown in Figure 1(c).

4.5 Comparative Summarization

Comparative document summarization aims to summarize the differences among comparable document groups. The summary produced for each group should emphasize its difference from other groups (Wang et al., 2009a).

We extend our method for update summarization to generate the discriminant summary for each group of documents. Given N groups of documents $C_1, C_2, \ldots, C_N$, we first generate the sentence graphs $G_1, G_2, \ldots, G_N$, respectively. To generate the summary for $C_i, 1 \leq i \leq N$, we view C_i as the update of all other groups. To extract a new sentence, only the one connected with the largest number of sentences which have no representatives in any groups will be extracted. We denote the extracted set as the complementary dominating set, since for each group we obtain a subset of vertices dominating those are not dominated by the dominating sets of other groups. To perform comparative summarization, we first extract the standard dominating sets for $G_1, \ldots, G_N$, respectively, denoted as $D_1, \ldots, D_N$. Then we extract the so-called complementary dominating set CD_i for G_i by continuing adding vertices in G_i to find the dominating set of $\cup_{1\leq j\leq N} G_j$ given $D_1, \ldots, D_{i-1}, D_{i+1}, \ldots, D_N$. A graphical illustration of comparative summarization is shown in Figure 1(d).

	DUC04	DUC05	DUC06	TAC08 A	TAC08 B
Type of Summarization	Generic	Topic-focused	Topic-focused	Topic-focused	Update
#topics	NA	50	50	48	48
#documents per topic	10	25-50	25	10	10
Summary length	665 bytes	250 words	250 words	100 words	100 words

Table 1: Brief description of the data set

5 Experiments

We have conducted experiments on all four summarization tasks and our proposed methods based on the minimum dominating set have outperformed many existing methods. For the generic, topic-focused and update summarization tasks, the experiments are perform the DUC data sets using ROUGE-2 and ROUGE-SU (Lin and Hovy, 2003) as evaluation measures. For comparative summarization, a case study as in (Wang et al., 2009a) is performed. Table 1 shows the characteristics of the data sets. We use DUC04 data set to evaluate our method for generic summarization task and DUC05 and DUC06 data sets for query-focused summarization task. The data set for update summarization, (i.e. the main task of TAC 2008 summarization track) consists of 48 topics and 20 newswire articles for each topic. The 20 articles are grouped into two clusters. The task requires to produce 2 summaries, including the initial summary (TAC08 A) which is standard query-focused summarization and the update summary (TAC08 B) under the assumption that the reader has already read the first 10 documents.

We apply a 5-fold cross-validation procedure to choose the threshold λ used for generating the sentence graph in our method.

5.1 Generic Summarization

We implement the following widely used or recent published methods for generic summarization as the baseline systems to compare with our proposed method (denoted as MDS). (1) Centroid: The method applies MEAD algorithm (Radev et al., 2004) to extract sentences according to the following three parameters: centroid value, positional value, and first-sentence overlap. (2) LexPageRank: The method first constructs a sentence connectivity graph based on cosine similarity and then selects important sentences based on the concept of eigenvector centrality (Erkan and Radev, 2004). (3) BSTM: A Bayesian sentence-based topic model making use of both the term-document and term-sentence associations (Wang et al., 2009b).

Our method outperforms the simple Centroid method and another graph-based LexPageRank, and its performance is close to the results of the Bayesian sentence-based topic model and those of the best team in the DUC competition. Note however that, like clustering or topic based methods, BSTM needs the topic number as the input, which usually varies by different summarization tasks and is hard to estimate.

5.2 Query-Focused Summarization

We compare our method (denoted as MWDS) described in Section 4.3 with some recently published systems. (1) TMR (Tang et al., 2009): incorporates the query information into the topic model, and uses topic based score and term frequency to estimate the importance of the sentences. (2) SNMF (Wang et al., 2008): calculates sentence-sentence similarities by sentence-level semantic analysis, clusters the sentences via symmetric nonnegative matrix factorization, and extracts the sentences based on the clustering result. (3) Wiki (Nastase, 2008): uses Wikipedia as external knowledge to expand query and builds the connection between the query and the sentences in documents.

Table 3 presents the experimental comparison of query-focused summarization on the two datasets. From Table 3, we observe that our method is comparable with these systems. This is due to the good interpretation of the summary extracted by our method, an ap-

	DUC04	
	ROUGE-2	ROUGE-SU
DUC Best	0.09216	0.13233
Centroid	0.07379	0.12511
LexPageRank	0.08572	0.13097
BSTM	0.09010	0.13218
MDS	0.08934	0.13137

Table 2: Results on generic summarization.

	DUC05		DUC06	
	ROUGE-2	ROUGE-SU	ROUGE-2	ROUGE-SU
DUC Best	0.0725	0.1316	0.09510	0.15470
SNMF	0.06043	0.12298	0.08549	0.13981
TMR	0.07147	0.13038	0.09132	0.15037
Wiki	0.07074	0.13002	0.08091	0.14022
MWDS	0.07311	0.13061	0.09296	0.14797

Table 3: Results on query-focused summarization.

proximate minimal dominating set of the sentence graph. On DUC05, our method achieves the best result; and on DUC06, our method outperforms all other systems except the best team in DUC. Note that our method based on the minimum dominating set is much simpler than other systems. Our method only depends on the distance to the query and has only one parameter (i.e., the threshold λ in generating the sentence graph).

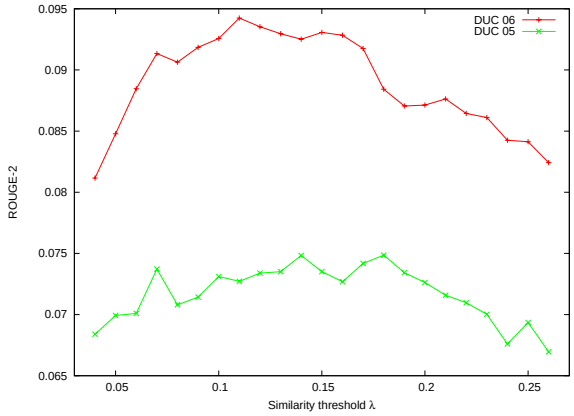

Figure 2: ROUGE-2 vs. threshold λ

We also conduct experiments to empirically evaluate the sensitivity of the threshold λ. Figure 2 shows the ROUGE-2 curve of our MWDS method on the two datasets when λ varies from 0.04 to 0.26. When λ is small, edges fail to represent the similarity of the sentences, while if λ is too large, the graph will be sparse. As λ is approximately in the range of $0.1 - 0.17$, ROUGE-2 value becomes stable and relatively high.

5.3 Update Summarization

Table 5 presents the experimental results on update summarization. In Table 5, 'TAC Best" and "TAC Median" represent the best and median results from the participants of TAC 2008 summarization track in the two tasks respectively according to the TAC 2008 report (Dang and Owczarzak, 2008). As seen from the results, the ROUGE scores of our methods are higher than the median results. The good results of the best team typically come from the fact that they utilize advanced natural language processing (NLP) techniques to resolve pronouns and other anaphoric expressions. Although we can spend more efforts on the preprocessing or language processing step, our goal here is to demonstrate the effectiveness of formalizing the update summarization problem using the minimum dominating set and hence we do not utilize advanced NLP techniques for preprocessing. The experimental results demonstrate that our simple update summarization method based on the minimum dominating set can lead to competitive performance for update summarization.

	TAC08 A		TAC08 B	
	ROUGE-2	ROUGE-SU	ROUGE-2	ROUGE-SU
TAC Best	0.1114	0.14298	0.10108	0.13669
TAC Median	0.08123	0.11975	0.06927	0.11046
MWDS	0.09012	0.12094	0.08117	0.11728

Table 5: Results on update summarization.

5.4 Comparative Summarization

We use the top six largest clusters of documents from TDT2 corpora to compare the summary generated by different comparative summarization methods. The topics of the six document clusters are as follows: topic 1: Iraq Issues; topic 2: Asia's economic crisis; topic 3: Lewinsky scandal; topic 4: Nagano Olympic Games; topic 5: Nuclear Issues in Indian and Pakistan; and topic 6: Jakarta Riot. From each of the topics, 30 documents are extracted

Topic	Complementary Dominating Set	Discriminative Sentence Selection	Dominating Set
1	··· U.S. Secretary of State Madeleine Albright arrives to consult on the stand-off between the **United Nations and Iraq**.	**the U.S. envoy to the United Nations**, Bill Richardson, ··· play down China's refusal to support threats of **military force against Iraq**	The United States and Britain do not trust **President Saddam** and wants *cdots*warning of serious consequences if **Iraq** violates the accord.
2	**Thailand's currency**, the baht, **dropped through a key psychological level** of ··· amid a regional sell-off sparked by escalating **social unrest in Indonesia**.	Earlier, driven largely by **the declining yen**, **South Korea's stock market fell** by ···, while the **Nikkei 225 benchmark index dipped** below 15,000 in the morning ···	*In the fourth quarter, IBM Corp. earned $2.1 billion, up 3.4 percent from $2 billion a year earlier.*
3	··· attorneys representing **President Clinton and Monica Lewinsky.**	The following night **Isikoff** ···, where he directly followed the recitation of the top-10 list: "Top 10 **White House Jobs That Sound Dirty.**"	In Washington, Ken Starr's grand jury continued its investigation of the **Monica Lewinsky matter**.
4	Eight women and six men were named Saturday night as the first **U.S. Olympic Snowboard Team** as their sport gets set to make its debut in **Nagano, Japan**.	*this tunnel is finland's cross country version of tokyo's alpine ski dome, and olympic skiers flock from russia, ···, france and austria this past summer to work out the kinks* ···	If the skiers the **men's super-G** and the **women's downhill** on Saturday, they will be back on schedule.
5	**U.S. officials** have announced **sanctions** Washington will impose on **India and Pakistan** for conducting **nuclear tests**.	The **sanctions** would stop all foreign aid except for humanitarian purposes, **ban military sales to India** ···	And **Pakistan's prime minister** says his country will sign the **U.N.'s comprehensive ban on nuclear tests** if **India** does, too.
6	··· remain in force around **Jakarta**, and at the Parliament building where **thousands of students staged a sit-in** Tuesday ···.	"**President Suharto** has given much to his country over the past 30 years, raising **Indonesia's** standing in the world ···	*What were the students doing at the time you were there, and what was the reaction of the students to the troops?*

Table 4: A case study on comparative document summarization. Some unimportant words are skipped due to the space limit. The bold font is used to annotate the phrases that are highly related with the topics, and italic font is used to highlight the sentences that are not proper to be used in the summary.

randomly to produce a one-sentence summary. For comparison purpose, we extract the sentence with the maximal degree as the baseline. Note that the baseline can be thought as an approximation of the dominating set using only one sentence. Table 4 shows the summaries generated by our method (complementary dominating set (CDS)), discriminative sentence selection (DSS) (Wang et al., 2009a) and the baseline method. Our CDS method can extract discriminative sentences for all the topics. DSS can extract discriminative sentences for all the topics except topic 4. Note that the sentence extracted by DSS for topic 4 may be discriminative from other topics, but it is deviated from the topic Nagano Olympic Games. In addition, DSS tends to select long sentences which should not be preferred for summarization purpose. The baseline method may extract some general sentences, such as the sentence for topic 2 and topic 6 in Table 4.

6 Conclusion

In this paper, we propose a framework to model the multi-document summarization using the minimum dominating set and show that many well-known summarization tasks can be formulated using the proposed framework. The proposed framework leads to simple yet effective summarization methods. Experimental results show that our proposed methods achieve good performance on several multi-document document tasks.

7 Acknowledgements

This work is supported by NSF grants IIS-0549280 and HRD-0833093.

References

Chen, Y.P. and A.L. Liestman. 2002. Approximating minimum size weakly-connected dominating sets for clustering mobile ad hoc networks. In *Proceedings of International Symposium on Mobile Ad hoc Networking & Computing*. ACM.

Chvatal, V. 1979. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233–235.

Dang, H.T. and K Owczarzak. 2008. Overview of the TAC 2008 Update Summarization Task. In *Proceedings of the Text Analysis Conference (TAC)*.

Dang, H.T. 2007. Overview of DUC 2007. In *Document Understanding Conference*.

Daumé III, H. and D. Marcu. 2006. Bayesian query-focused summarization. In *Proceedings of the ACL-COLING*.

Erkan, G. and D.R. Radev. 2004. Lexpagerank: Prestige in multi-document text summarization. In *Proceedings of EMNLP*.

Feige, U. 1998. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652.

Goldstein, J., V. Mittal, J. Carbonell, and M. Kantrowitz. 2000. Multi-document summarization by sentence extraction. In *NAACL-ANLP 2000 Workshop on Automatic summarization*.

Guha, S. and S. Khuller. 1998. Approximation algorithms for connected dominating sets. *Algorithmica*, 20(4):374–387.

Haghighi, A. and L. Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of HLT-NAACL*.

Han, B. and W. Jia. 2007. Clustering wireless ad hoc networks with weakly connected dominating set. *Journal of Parallel and Distributed Computing*, 67(6):727–737.

Johnson, D.S. 1973. Approximation algorithms for combinatorial problems. In *Proceedings of STOC*.

Jurafsky, D. and J.H. Martin. 2008. *Speech and language processing*. Prentice Hall New York.

Kann, V. 1992. *On the approximability of NP-complete optimization problems*. PhD thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm.

Lawrie, D., W.B. Croft, and A. Rosenberg. 2001. Finding topic words for hierarchical summarization. In *Proceedings of SIGIR*.

Lin, C.Y. and E. Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of HLT-NAACL*.

Mani, I. 2001. Automatic summarization. *Computational Linguistics*, 28(2).

Nastase, V. 2008. Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation. In *Proceedings of EMNLP*.

Nenkova, A. and L. Vanderwende. 2005. The impact of frequency on summarization. *Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101*.

Radev, D.R., H. Jing, M. Styś, and D. Tam. 2004. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40(6):919–938.

Raz, R. and S. Safra. 1997. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of STOC*.

Saggion, H., K. Bontcheva, and H. Cunningham. 2003. Robust generic and query-based summarisation. In *Proceedings of EACL*.

Tang, J., L. Yao, and D. Chen. 2009. Multi-topic based Query-oriented Summarization. In *Proceedings of SDM*.

Thai, M.T., N. Zhang, R. Tiwari, and X. Xu. 2007. On approximation algorithms of k-connected m-dominating sets in disk graphs. *Theoretical Computer Science*, 385(1-3):49–59.

Wan, X., J. Yang, and J. Xiao. 2007a. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of IJCAI*.

Wan, X., J. Yang, and J. Xiao. 2007b. Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction. In *Proceedings of ACL*.

Wang, D., T. Li, S. Zhu, and C. Ding. 2008. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of SIGIR*.

Wang, D., S. Zhu, T. Li, and Y. Gong. 2009a. Comparative document summarization via discriminative sentence selection. In *Proceeding of CIKM*.

Wang, D., S. Zhu, T. Li, and Y. Gong. 2009b. Multi-document summarization using sentence-based topic models. In *Proceedings of the ACL-IJCNLP*.

Wei, F., W. Li, Q. Lu, and Y. He. 2008. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *Proceedings of SIGIR*.

Wu, J. and H. Li. 2001. A dominating-set-based routing scheme in ad hoc wireless networks. *Telecommunication Systems*, 18(1):13–36.