Query-Focused Multi-Document Summarization Using Co-Training Based Semi-Supervised Learning *

Po Hu [a,b], Donghong Ji [a], Hai Wang [c], and Chong Teng [a]

[a] 129 Luoyu Road, Computer School, Wuhan University,
Wuhan 430079, China
phu@mail.ccnu.edu.cn, donghong_ji2000@yahoo.com.cn, tchong616@126.com
[b] 152 Luoyu Road, Department of Computer Science, Huazhong Normal University,
Wuhan 430079, China
[c] 463 Guanshan Road, School of Electronics and Information, Wuhan Polytechnic,
Wuhan 430074, China
whseaking@163.com

Abstract. This paper presents a novel approach to query-focused multi-document summarization. As a good biased summary is expected to keep a balance among query relevance, content salience and information diversity, the approach first makes use of both the content feature and the relationship feature to select a number of sentences via the co-training based semi-supervised learning, which can identify the query relevant sentences beyond a single point of view. Then the ranking algorithm based on Markov chain random walks is employed on the relevant sentences by encouraging content salience and information diversity in a unified framework. The final summary focusing on the integration of relevance, salience and diversity is created after several sentences with the highest overall ranking scores are extracted. We performed experiments on DUC2007 dataset and the evaluation results show that the proposed approach can achieve significant improvement over standard baseline approaches and gain comparable performance to the state-of-the-art systems.

Keywords: multi-document summarization, co-training based semi-supervised learning, graph based sentence ranking

1 Introduction

Query-focused multi-document summarization has attracted much attention in recent years. Different from generic summarization, it aims to provide more personalized information for a given query. By automatically capturing relevant and salient content from a large amount of searching results and showing them in a concise way, the sort of summarization can aid people to quickly access and digest their interested information. It also provides an effective means to diverse applications such as question answering system, personalized information retrieval, personalized news recommender, etc. To date, the most influential annual evaluation workshop for automatic summarization research is the Document Understanding Conference (DUC or now TAC), which provides a large-scale test benchmark as well as common evaluation procedures for researchers to share their ideas and experiences.

* The work reported in this paper was supported by the Major Research Plan of National Natural Science Foundation of China (90820005, 90920005), National Natural Science Foundation of China (60773011, 60773167) and Wuhan University 985 Project (985yk004).

The critical issues in query-focused multi-document summarization are as follows: The first one is that the information contained in the generated summary should be highly related to the given query. The second issue is how to take salience and diversity into account when selecting a batch of sentences or other textual units as representatives for a summary. As the allowed capacity in a summary is usually limited, it will be inappropriate to put all the informative sentences from different documents into the summary for they may convey the similar meanings. The intuition behind a good query-focused summary is to preserve the information biased to the query as much as possible, and remain the most representative and salient information that have the least duplicate contents to the information selected previously.

In this study, we propose a novel sentence-based extractive approach. Since it is not enough to select relevant sentences from a single point of view and no labeled relevant or irrelevant sentences are available in advance, the proposed approach first makes full use of the co-training algorithm to identify the relevant sentences on two abundant feature views with a small number of pseudo-labeled sentences. Then it employs a ranking algorithm to sort the relevant sentences and choose a certain number of representatives with highest content salience and information diversity for the final summary. Experimental results on DUC2007 main task dataset show that the proposed approach significantly outperforms the baseline approaches and achieves comparable performance to the state-of-the-art systems.

The remainder of the paper is organized as follows. Section 2 discusses related work. The proposed summarization approach is described in Section 3. The details of the experimentation are shown in Section 4. Section 5 presents our conclusion and future work.

2 Related Work

Most of multi-document summarization methods can be categorized into two main paradigms, i.e. extractive and abstractive summarization. Extractive summarization often directly extracts important sentences in supervised, unsupervised or semi-supervised way based on the combination of a few implicit or explicit features (Goldstein et al., 2000; Radev et al., 2004), while abstractive summarization usually makes use of deep natural language understanding or generation technology to fuse or reformulate information (Knight and Marcu, 2000). In this paper, we focus on extractive summarization approach.

Compared to single-document summarization, it is more likely for multi-document summarization to have repetitive contents and diverse subtopics across documents, so maximizing content salience and minimizing content redundancy has been recognized as one of the major difficulties. Many methods have been proposed to achieve this goal. Maximum marginal relevance (MMR) (Carbonell and Goldstein, 1998), GRASSHOPPER ranking (Zhu et al., 2007), diversity penalty (Zhang et al., 2005) and mixture models (Zhang et al., 2002) are commonly used approaches incorporating information salience and diversity into the ranking process. Among these approaches, GRASSHOPPER ranking tries to encourage the balance of salience and diversity in a unified framework, while other approaches deal with them separately.

Inspired by PageRank and HITS algorithm, much focus has been put on adopting graph-based ranking algorithm like LexRank (Erkan and Radev, 2004) and TextRank (Mihalcea and Tarau, 2004) to multi-document summarization. These algorithms generally employ the global information described by a passage affinity graph and recursively calculate each passage's significance based on link structure analysis, stability-based random walk, global consistency or smoothness-based label propagation on the graph. Topic-sensitive LexRank (Haveliwala, 2003) extended the traditional LexRank algorithm by integrating the similarity between sentences and the given query. Wan et al. (2007) adopted a manifold-ranking algorithm to rank sentences by considering global information and emphasizing the high biased information richness in a score propagation process.

Recently co-training algorithm has been successfully used in many natural language processing applications (Muller et al., 2002; Sarkar, 2001). Wong et al. (2008) applied co-

training algorithm to generic multi-document summarization, which trains two different classifiers on the same feature space to evaluate the importance of a sentence and needs a few manually labeled examples as training data. However, there is little research in applying co-training based learning algorithm to query-focused multi-document summarization especially when manually labeled information is absent. In our approach, the major point of concern is how to employ the co-training algorithm to support better choosing query relevant sentences from two different but abundant feature views, which can incorporate multi-dimensional complementary information to classify each sentence by leveraging both the individual information in each sentence and the relationship information among sentences.

3 The Proposed Approach

3.1 Overview

In order to generate a biased summary with good characteristics of relevance, salience and diversity, we first investigate the effectiveness of combining two kinds of different features in a semi-supervised learning process to decide which sentences are relevant to the query. After that, a sentence ranking algorithm with the emphasis on salience and diversity is employed on these relevant sentences so as to ensure the top-ranked sentences such characteristics.

The proposed approach mainly consists of two steps. The first step is to classify all the sentences in a document set under two view settings and select a number of sentences closely related to the query via the co-training based learning algorithm (Blum and Mitchell, 1998). The second step is to rank all the selected sentences via the Markov chain random walks (Zhu et al., 2007) and take the top ones to create the final summary, which takes into account content salience and information diversity in a unified framework.

3.2 Co-training Based Learning for the Selection of Query Relevant Sentences

In query-focused multi-document summarization, a user usually poses a query reflecting his personalized requirement and asks the summarizer to answer the query in a concise way. So how to find the relevant sentences becomes the primary problem. Semi-supervised learning is a natural approach for this problem where all the sentences in a document set are required to be classified into positive sentences (i.e. query relevant sentences) and negative sentences (i.e. query irrelevant sentences), but only one positive labeled sentence (i.e. the query) is available.

Co-training is a classic semi-supervised learning algorithm that can take advantage of unlabeled data to boost learning performance. Traditional co-training works under a two-view setting and assumes that each example should be described by two different and conditionally independent feature sets. Recent research has demonstrated that the strong assumption of independence between two views is not necessary. However co-training can not be effectively performed when there is only one labeled example, so our approach tries to choose a small number of sentences as pseudo labeled sentences. For pseudo relevant sentences, we determine a few sentences with the highest similarity to the query; for pseudo irrelevant ones, we pick a few sentences with the lowest similarity to the query. In this paper, we focus on how to make use of two sufficient features in the co-training process to classify the sentences and automatically infer the labels for the unlabeled sentences.

Given a set of sentences $S=\{s_0, s_1,\ldots,s_{n-1}\}$. Here s_0 denotes the given query that can be treated as a pseudo-sentence, and the rest denotes all the sentences in the documents to be summarized. Let X and Y denote two different sentence features investigated in our approach. Here X represents content feature that uses content bearing terms to describe a sentence, and Y is relationship feature that represents a sentence by its pair-wise similarity with other sentences. We use matrix $[M_{ij}]_{n*m}$ to describe the sentence set that is formally represented on feature X with each entry M_{ij} corresponding to the weight associated with term t_j in sentence s_i, which is calculated by the $TF_{ij}*ISF_i$ formula, where n is the total number of sentences including the

query, m is the total number of terms in the documents, TF_{ij} denotes the frequency of term t_j appearing in sentence s_i, and ISF_i is the inverse sentence frequency of term t_j, which is calculated by $1+\log(n/n_j)$, where n_j is the number of the sentences that contain term t_j. So each sentence can be represented by an m-dimensional term vector. We also use another matrix $[N_{ij}]_{n*n}$ to describe the sentences on feature Y with each entry N_{ij} corresponding to the similarity between sentence s_i and s_j, which can reflect the pair-wise relationship. Here $[N_{ij}]_{n*n}$ is calculated by formula 1,

$$N_{ij}=\begin{cases}\dfrac{\overrightarrow{V_{s_i}}\bullet\overrightarrow{V_{s_j}}}{\|\overrightarrow{V_{s_i}}\|\times\|\overrightarrow{V_{s_j}}\|} & (i\neq j)\\ 1 & (i=j)\end{cases} \quad (1)$$

where $\overrightarrow{V_{s_i}}, \overrightarrow{V_{s_j}}$ are the corresponding term vectors for sentence s_i and s_j.

When a certain number of pseudo-labeled sentences are available[1], the next task is to classify the rest unlabeled sentences into positive and negative ones. Table 1 gives the co-training based learning algorithm for the selection of query relevant sentences.

Table 1: The co-training based learning algorithm for the selection of query relevant sentences.

Input:

Matrix $[M_{ij}]_{n*m}$ and $[N_{ij}]_{n*n}$, which denote all the sentences' formal representations on content feature view X and relationship feature view Y respectively.

L is the set of pseudo-labeled relevant and irrelevant sentences.

U is the set of unlabeled sentences.

Output:

The predicted labels for all the unlabeled sentences in the documents and a number of selected query relevant sentences with the positive label.

Process:

1. Establish the mapping between the sentences in L, U and the row vectors in $[M_{ij}]_{n*m}$ and $[N_{ij}]_{n*n}$.
2. Create an unlabeled sentence pool U' by selecting u sentences from U at random.
3. Loop while there are still some unlabeled sentences in U
   - Use L to train a classifier C_x on $[M_{ij}]_{n*m}$.
   - Use L to Train a classifier C_y on $[N_{ij}]_{n*n}$.
   - Use C_x to label p positive and n negative sentences with the highest classifying confidence from U'.
   - Use C_y to label p positive and n negative sentences with the highest classifying confidence from U'.
   - Add these labeled sentences to L and remove them from U.
   - Randomly choose 2(p+n) sentences from U to replenish U'.

The above algorithm is intuitively based on the following assumptions:

Assumption 1: A sentence should be highly related to the query if it contains the same or similar content bearing terms in the query.

Assumption 2: A sentence should be highly related to the query if it has the same or similar relationship distribution with other sentences like the query.

Assumption 3: co-training can exploit the richer information described by two different features to train classifiers so as to select the relevant sentences from unlabeled sentences.

[1] The given query is regarded as a labeled relevant (i.e. positive) sentence in our approach.

3.3 Random Walks Based Ranking for the Selection of Salient and Diverse Sentences

The GRASSHOPPER ranking algorithm is a general-purpose ranking method, which focuses on information diversity during the ranking process. The underlying idea of the algorithm is that the items and inter-item relationships can be encoded by a graph. We can define a random walk on the graph correspondingly and determine the importance of a node (i.e. item) by stationary distribution of random walk on the graph. If a node is most similar to many other nodes, it will first become a highly ranked one and at the same time be adjusted into the absorbing state, which will cut down the significance of similar unranked nodes and encourage diversity. In our study, the GRASSHOPPER ranking algorithm (Zhu et al., 2007) is applied to achieve both content salience and information diversity in a unified framework. The whole procedure of the random walks based ranking algorithm goes as follows:

---

1. Construct an undirected affinity graph G_r over the query relevant sentences that have been selected in the co-training process, where each sentence is considered as a node and edges are created between two sentences if their pair-wise similarity exceeds 0.01.
2. Define an adjacency matrix M_r to represent G_r with each entry corresponding to the cosine similarity of two corresponding sentence vectors.
3. Normalize matrix M_r to matrix $\widetilde{M}_r$ by dividing each element in M_r by the corresponding row sum.
4. Use $\widetilde{M}_r$ to form a stochastic matrix M_s by integrating a prior ranking distribution r on these sentences according to formula 2.

$$M_s = \lambda \widetilde{M}_r + (1-\lambda) 1 r^T \quad (2)$$

M_s can be considered as the transition matrix of a Markov chain with the entry $M_s(i,j)$ specifying the transition probability from state i (i.e. sentence s_i) to state j (i.e. sentence s_j) in the corresponding Markov chain. $\lambda \in [0,1]$ is a damping factor, 1 is an all-1 vector, and $1r^T$ denotes the prior ranking that is represented as a probability distribution. The teleporting random walks based on M_s act in such a way that moving to an adjacent state according to the entry in $\widetilde{M}_r$ with probability λ or jumping to a random state according to the prior ranking distribution with probability $1-\lambda$ at each step.

5. Compute M_s's stationary distribution and take the sentence (i.e. state) with the largest stationary probability to be the top one for the final ranking.
6. Turn ranked sentences into absorbing states and compute the expected number of visits for all the rest sentences. Then pick the next higher ranked sentence with the maximum expected number of visits[2]. Repeat step 6 until all the relevant sentences are ranked.

In the above algorithm, the sixth step is crucial for encouraging diversity in ranking because it prefers those sentences that have more salient visit opportunities and prevents those that highly related to the higher ranked sentences from getting high score for the random walks are apt to be absorbed soon after visiting them. After the above ranking process, a number of query relevant sentences with high content salience and information diversity are extracted and concatenated to create the final summary in accordance with the length limit.

[2] An illustration that the Markov chain with the stochastic matrix M_s will converge to a unique stationary distribution and the detailed description about how to compute the expected number of visits in an absorbed Markov chain can be found at the reference papers written by Zhu et al., 2007.

4 Experiments

4.1 Data Set

To evaluate the effectiveness of the proposed approach, we experiment on DUC2007 main task dataset. The dataset consists of 45 topics with each topic comprising a query description as topic narrative, a set of 25 relevant documents from the AQUAINT corpus and 4 human model summaries for evaluation. The main task in DUC2007 is to create from the document set a brief, well-organized, fluent summary with its length no longer than 250 words which answers the need for information expressed in the query.

4.2 Evaluation Metric

The ROUGE toolkit (Lin and Hovy, 2003), which is the most frequently used automated summary evaluation package in annual DUC and TAC, is adopted for evaluating our experiment results. Usually this toolkit is employed to measure how much of the contents of a set of human-produced "standard" summaries are contained by an automatically produced summary. By automatically comparing various levels of content unit's overlap between the automatically created summary and the reference manual summaries, a few recall-oriented ROUGE metrics have been proposed such as ROUGE-1 (unigram based metric), ROUGE-2 (bigram based metric) and ROUGE-SU4(skip bigram and unigram based metric with maximum skip distance 4), etc. Among them, ROUGE-1 metric has been shown to correlate with the human judgment best. We present the metric scores of ROUGE-1, ROUGE-2 and ROUGE-SU4 at the confidence level of 95% in the following experiment for they have been officially used in DUC2007 for system's comparison.

4.3 Experimental Results

In the following experiments, queries and documents were segmented into sentences, stop-words were removed and the remaining words were stemmed by Porter Stemmer. All the sentences and the queries were represented as the term vectors according to TF*ISF scheme. The relevance of a sentence to the query and other sentences were computed by cosine similarity on their corresponding term vectors. The proposed approach was first compared with two baseline approaches and other systems participating in DUC2007 main task.

Baseline 1: A simple summarizer that returns the first 250 words of the most recent document in the topic.

Baseline 2: A generic multi-document summarizer named CLASSY04, which ignores the query narrative information but has the highest evaluation score in Task 2 of DUC 2004.

In our approach, the numbers of pseudo relevant and irrelevant sentences, which are determined by the highest and lowest similarity to the query respectively, are set to 25 empirically. The selection of query relevant sentences adopts the co-training based algorithm described in Section 3.2. The parameters of the co-training algorithm are set as follows: p and n are set to 1, which denote the number of positive and negative candidates in each iteration, u represents the pool size denoting the number of sentences selected from the set of unlabeled sentences randomly, which is set to 75, J48 (known as C4.5) decision tree classifier implemented in Weka (Witten and Frank, 2005) is used to train the classifiers C_x and C_y, and the classifying confidence threshold is set to 0.95. The damping factor λ of the Markov chain random walks based ranking (known as GRASSHOPPER) is set to 0.9, and we let the prior ranking $1r^T$ the uniform probability distribution vector because there is no explicit prior ranking in our co-training based selection process. Table 2 shows the comparison results.

Table 2: System comparison results

	ROUGE-1	ROUGE-2	ROUGE-SU4
Baseline 1	0.31250	0.06039	0.10507
Baseline 2	0.40562	0.09382	0.14641
Best/worst performance of participating systems	0.45258/ 0.24277	0.12448/ 0.03813	0.17711/ 0.07385
Mean performance of participating systems	0.39728	0.09486	0.14747
Our approach	**0.42298**	**0.10824**	**0.16131**

The ROUGE evaluation results in Table 2 show that the proposed approach can achieve significant improvement over baseline approaches and gain comparable performance to the state-of-the-art systems. The encouraging performance achieved by the proposed approach can be attributed to the following factors.

1) Co-training based selection of query relevant sentences

The Co-training based algorithm can make full use of both the content feature and the relationship feature to learn to classify sentences into two classes (i.e. query relevant sentences and query irrelevant sentences). It also allows a large amount of unlabeled sentences to augment a much smaller set of pseudo-labeled sentences.

2) GRASSHOPPER based selection of salient and diverse sentences

GRASSHOPPER ranking is an alternative to MMR and tries to achieve salience and diversity in a unified framework. It can make use of the random walks in an absorbing Markov chain to rank a set of relevant sentences so that the highly ranked sentences have higher local centrality and cover as many distinct subtopics as possible.

4.4 Effect of the Size of the Pseudo Relevant and Irrelevant Sentences

To study the influence of the number of the pseudo relevant and irrelevant sentences in the propose approach, we repeat our experiments with different size of pseudo labeled sentences which varies from 5 to 75 with the step set to 10. Due to page limit, we only report ROUGE-1 performance as representative in this experiment. Figure 1 demonstrates the effect of the parameter while other parameters remain unchanged.

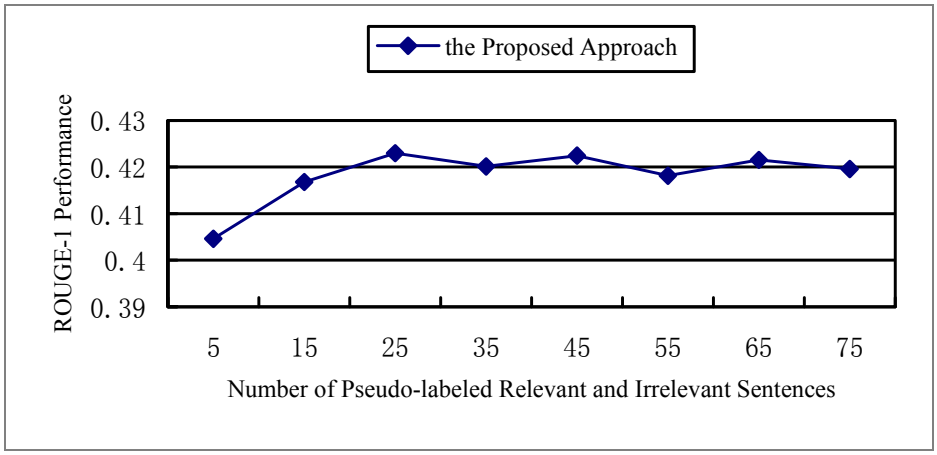

Figure 1: ROUGE-1 performance vs. number of pseudo-labeled relevant and irrelevant sentences.

From Figure 1, we can find that the proposed approach can achieve 0.405 on ROUGE-1 metric when only five pseudo-labeled sentences are available, which is better than the average performance of the participating systems in DUC2007 main task and suggests that a small

number of pseudo-labeled samples are enough for the approach. The ROUGE-1 score increases sharply when the number of pseudo-labeled sentences varies from 5 to 25 and achieves the summit at 25. Then it tends to be gradually stabilized at 0.42 around, which doesn't show a significant difference as the number of pseudo-labeled sentences increases. One possible explanation for the tendency is that more pseudo-labeled relevant and irrelevant sentences provided by the similarity based selection strategy will also bring more noise that will influence the co-training performance afterward in a certain degree. This also verifies that the proposed co-training based summarization method can achieve promising and robust performance when an appropriate amount of pseudo-labeled sentences are provided.

4.5 Effect of the Selection Strategies for Query Relevant Sentences

To explore the impact of different selection strategies for relevant sentences on the evaluation result, another approach (i.e. relevance-based approach) is implemented as reference, which first computes the similarity-based relevance between the query and each sentence in the document set and chooses those sentences whose relevance to the query exceeds zero as candidates. Then GRASSHOPPER ranking algorithm is imposed on the candidate sentence set and the sentences with highest ranking scores are selected to create the final summary.

In relevance-based approach the damping factor λ of the GRASSHOPPER ranking is set to 0.5 because an explicit prior ranking (i.e. the similarity relevance) for each relevant sentence can be utilized and it may contribute to the final ranking score equally as another factor encoded by the affinity graph does.

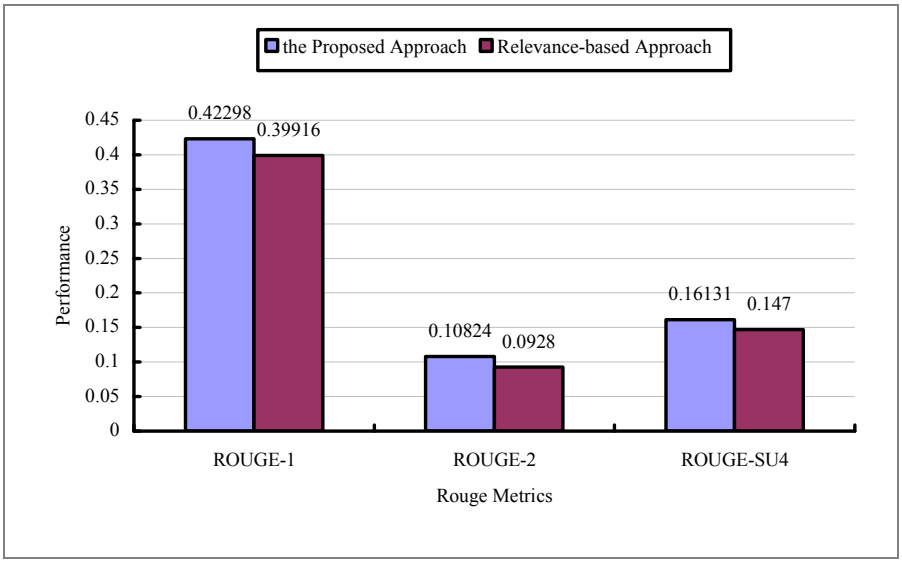

Figure 2: ROUGE performance comparison between the proposed approach and relevance-based approach.

From Figure 2, it can be found that the co-training based algorithm is better than the purely relevance-based strategy when selecting a number of query relevant sentences. It indicates that combining the co-trained information from both content feature and relationship feature can provide better heuristics for the selection process and lead to better performance.

4.6 Effect of the Selection Strategies for Salient and Diverse Sentences

If the proposed approach does not conduct the random walks based ranking (i.e. let λ=0 in GRASSHOPPER ranking) on those selected relevant sentences, we observe that the ROUGE-1 score will decrease dramatically from 0.42298 to 0.39268 because many sentences with high

query relevance will be selected as the summary sentences but they may convey the similar or even the same information about the same topic. The result clearly shows that if the proposed approach does not take into account the content salience and information diversity among the relevant sentences, it will deteriorate the final performance evidently. However, when the prior ranking of relevant sentences is not considered (i.e. let λ=1 in GRASSHOPPER ranking), the performances of both approaches are still well.

5 Conclusion and Future Work

This paper proposes a novel extractive approach to query-focused multi-document summarization. The proposed approach can make full use of the co-training based semi-supervised learning algorithm to identify the relevant sentences on two abundant feature views with a small number of pseudo-labeled sentences. The final summary created by the approach can keep a balance among query relevance, content salience and information diversity in a unified framework. Experimental results on DUC2007 main task dataset demonstrate the effectiveness of the proposed approach and verify the potential of the co-training based learning algorithm in query-focused summarization.

In future work, we will apply the co-training base method to other summarization tasks such as update summarization, opinion summarization, etc. Moreover, to show the advantage of using co-training in feature combination, we will make a comparison to traditional combination approaches such as linear combination and multiple modalities fusion. Future work also includes exploring how to use other different feature views to extract query relevant sentences and make the final summary more informative, coherent and readable.

References

Blum, A. and T. Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT'98)*, pp.92-100.

Carbonell, J. and J. Goldstein 1998. the Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'98)*, pp.335–336.

Christoph, M., S. Rapp and M. Strube. 2002. Applying Co-Training to Reference Resolution. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pp.352–359.

Erkan, G. and D.R. Radev. 2004. LexRank: Graph-Based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research*, 22:457-479.

Goldstein, J., V. Mittal, J. Carbonell, and J. Callan. 2000. Multi-Document Summarization by Sentence Extraction. *Proceedings of ANLP/NAACL Workshop on Summarization*, pp.40–48.

Haveliwala, T. H. 2003. Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 784-796.

Knight, K. and D. Marcu. 2000. Statistics-Based Summarization - Step One: Sentence Compression. *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence (AAAI'00)*, pp.703-710.

Lin C.Y. and H. Eduard. 2003. Automatic Evaluation of Summaries Using N-Gram Co-Occurrence Statistics. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'03)*, pp.71–78.

Mihalcea, R. and P. Tarau. 2004. TextRank: Bringing Order into Texts. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*, pp.404-411.

Muller, Christoph , Stefan Rapp and Michael Strube. 2002. Applying Co-Training to Reference Resolution. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*.

Radev, D.R., H.Y. Jing, M. Stys and D.Tam. 2004. Centroid-Based Summarization of Multiple Documents. *Information Processing and Management*, 40:919-938.

Sarkar, A. 2001. Applying Co-Training Methods to Statistical Parsing. *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'01)*, pp.175–182.

Wan, X.J., J.W. Yang and J.G. Xiao. 2007. Manifold-Ranking Based Topic-Focused Multi-Document Summarization. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 07)*, pp.2903-2908.

Witten, I.H. and E. Frank. 2005. *Data Mining:Practical Machine Learning Tools and Techniques,2nd Edition*, Morgan Kaufmann, San Francisco.

Wong, K.F., M.L. Wu and W.J. Li. 2008. Extractive Summarization Using Supervised and Semi-Supervised Learning. *Proceedings of the International Conference on Computational Linguistics (COLING'08)*, pp.985-992.

Zhang, B.Y., H. Li, Y. Liu, L. Ji, W.S. Xi, W.G. Fan, Z. Chen and W.Y. Ma. 2005. Improving Web Search Results Using Affinity Graph. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'05)*, pp.504–511.

Zhang, Y., J. Callan and T. Minka. 2002. Novelty and Redundancy Detection in Adaptive Filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR'02)*, pp.81–88.

Zhu, X.J., A. Goldberg, J.V. Gael and D. Andrzejewski. 2007. Improving Diversity in Ranking Using Absorbing Random Walks. *Proceedings of HLT-NAACL'07*, pp.97–104.