Using Syntactic and Shallow Semantic Kernels to Improve Multi-Modality Manifold-Ranking for Topic-Focused Multi-Document Summarization

Yllias Chali
University of Lethbridge
Lethbridge, AB, Canada
chali@cs.uleth.ca

Sadid A. Hasan
University of Lethbridge
Lethbridge, AB, Canada
hasan@cs.uleth.ca

Kaisar Imam
University of Lethbridge
Lethbridge, AB, Canada
imam@uleth.ca

Abstract

Multi-modality manifold-ranking is recently used successfully in topic-focused multi-document summarization. This approach is based on Bag-Of-Words (BOW) assumption where the pair-wise similarity values between sentences are computed using the standard cosine similarity measure (TF*IDF). However, the major limitation of the TF*IDF approach is that it only retains the frequency of the words and disregards the syntactic and semantic information. In this paper, we propose the use of syntactic and shallow semantic kernels for computing the relevance between the sentences. We argue that the addition of syntactic and semantic information can improve the performance of the multi-modality manifold-ranking algorithm. Extensive experiments on the DUC benchmark datasets prove the effectiveness of our approach.

1 Introduction

Text summarization is a good way to compress a huge amount of information into a concise form by selecting the most important information and discarding redundant information. According to Mani (2001), automatic text summarization takes a partially-structured source text from multiple texts written about the same topic, extracts information content from it, and presents the most important content to the user in a manner sensitive to the user's needs. In contrast to summarizing one document that is termed as single document summarization, multi-document summarization deals with multiple documents as sources that are related to one main topic under consideration. As compared to generic summarization that must contain the core information central to the source documents, the main goal of topic-focused multi-document summarization (i.e. query-based multi-document summarization) is to create from the documents a summary that can answer the need for information expressed in the topic or explain the topic (Wan et al., 2007). In this paper, we consider the problem of producing extraction-based[1] topic-focused multi-document summaries given a collection of documents.

In recent years, a variety of manifold-ranking based methods are applied successfully to topic-focused multi-document summarization. The basic manifold-ranking method is a typical graph-based summarization method that makes uniform use of the sentence-to-sentence relationships and the sentence-to-topic relationships in a manifold-ranking process (Wan et al., 2007). In the multi-modality manifold-ranking algorithm, sentence relationships are classified into within-document relationships and cross-document relationships, and each kind of relationships are considered as a separate modality (graph) (Wan and Xiao, 2009). These methods are based on Bag-Of-Words (BOW) assumption where the pair-wise similarity values between the sentences are computed using the standard cosine measure (TF*IDF). The major limitation of the TF*IDF approach is that it only retains the frequency of the words and does not take into account the sequence of them (word ordering). It ignores the syntactic and semantic structure of the sentences and thus, cannot distinguish between "The police shot the gunman" and "The gunman shot the police". Traditionally, information extraction techniques are based on the BOW approach augmented by language modeling. But when the task like *multi-document summarization* requires the use of more

[1] An extract summary consists of sentences extracted from the document while an abstract summary employs words and phrases not appearing in the original document (Mani and Maybury, 1999).

Proceedings of the 5th International Joint Conference on Natural Language Processing, pages 1098–1106,
Chiang Mai, Thailand, November 8 – 13, 2011.

complex semantics, the approaches based on only BOW are often inadequate to perform fine-level textual analysis. Although some improvements on BOW are given by the use of dependency trees and syntactic parse trees (Hirao et al., 2004), (Punyakanok et al., 2004), (Zhang and Lee, 2003b), but these too are not adequate in terms of documents having very long and articulated sentences or even paragraphs. Shallow semantic representations could prevent the sparseness of deep structural approaches and the weakness of BOW models (Moschitti et al., 2007). Thus, attempting an application of syntactic and semantic information in measuring the relevance between the sentences seems natural and hardly controversial.

In this paper, we extensively study the impact of syntactic and semantic information in computing the similarity between the sentences in the multi-modality manifold learning framework for topic-focused multi-document summarization. We believe that the augmentation of the similarity measures based on the syntactic and semantic information could be helpful to characterize the relation between the sentences in a more effective way than the traditional TF*IDF based similarity measures alone. To include syntactic and semantic information into the multi-modality manifold-ranking framework, we apply the tree kernel functions (Collins and Duffy, 2001) and re-implement the syntactic and shallow semantic tree kernel model according to Moschitti et al. (2007). We run our experiments on the DUC[2]-2006 benchmark dataset, and the results show that the addition of syntactic and semantic information improves the performance of the BOW-based multi-modality manifold-ranking approach. The rest of this paper is organized as follows: Section 2 focuses on the related work, Section 3 describes the multi-modality manifold ranking model, Section 4 discusses the syntactic and shallow semantic kernels, Section 5 presents the experimental details with evaluation results and finally, Section 6 concludes the paper.

2 Related Work

In recent years, researchers have become more interested in topic-focused summarization and hence, different methods have been proposed ranging from heuristic extensions of generic summarization schemes (by incorporating topic-biased information) to novel ones. For instance, Nastase (2008) expands the query by using encyclopedic knowledge in Wikipedia and use the topic expanded words with activated nodes in the graph to produce an extractive summary. Hal Daumé and Marcu (2006) present BAYESUM ("Bayesian summarization"), a sentence extraction model for query-focused summarization.

Wan et al. (2007) propose a manifold-ranking method to make uniform use of sentence-to-sentence and sentence-to-topic relationships whereas the use of multi-modality manifold-ranking algorithm is shown in Wan and Xiao (2009). However, these methods use the standard cosine similarity measure to compute the relatedness between the sentences ignoring the syntactic and semantic information. The importance of syntactic and semantic features in finding textual similarity is described by Zhang and Lee (2003a), Moschitti et al. (2007), and Moschitti and Basili (2006). An effective way to integrate syntactic and semantic structures in machine learning algorithms is the use of *tree kernel* functions (Collins and Duffy, 2001) which has been successfully applied to question classification (Zhang and Lee, 2003a; Moschitti and Basili, 2006). In this paper, we use the tree kernel functions and to the best of our knowledge, no study has used tree kernel functions before to encode syntactic/semantic information for more complex tasks such as computing the relatedness between the sentences in the multi-modality manifold ranking algorithm for topic-focused multi-document summarization.

3 Multi-Modality Manifold-Ranking Model

In this section, we present the theoretical details of the manifold-ranking method (Zhou et al., 2003a; Zhou et al., 2003b), a universal ranking algorithm. This method is employed to rank data points and has been successfully applied in topic-focused document summarization in Wan et al. (2007) where the data points refer to the topic description and all the sentences in the documents. The manifold-ranking process for the summarization task can be formalized as follows (Wan and Xiao, 2009):

Given a set of data points , the first point represents the topic description (query point) and

[2]http://duc.nist.gov/

the rest points represent all the sentences in the documents (data points to be ranked). The basic manifold-ranking algorithm treats the sentence relationships in a single modality (Wan et al., 2007) whereas, in Wan and Xiao (2009), the relationships between the sentences in a document set are classified as either within-document relationship or cross-document relationship to form two separate modalities to reflect the local information channel and the global information channel between the sentences, respectively. The two modalities are applied in the multi-modality manifold-ranking algorithm for ranking the sentences effectively. Based on each kind of modality, an undirected graph is built to reflect each kind of sentence relationships. Let be the within-document affinity matrix containing only the within-document links for the data points, where is the cosine similarity[3] value between and if and belong to the same document or one of and is ; Otherwise, is set to . Similarly, let be the cross-document affinity matrix containing the cross-document links, where is the cosine similarity value between and if and belong to different documents or one of and is ; Otherwise, is set to . All the relationships between the topic, and any document sentence are included in both and . Then, and are normalized by ¯ ¯ and ¯ ¯, respectively, where and are the diagonal matrices with -element equal to the sum of the th row of and , respectively. Then the multi-modality learning task for topic-focused summarization is to infer the ranking function from , and : .

Linear Fusion: For fusing the two modalities, we use the linear fusion scheme as this was shown to perform the best in Wan and Xiao (2009). This scheme fuses the constraints from , and simultaneously by a weighted sum. The cost function associated with is defined as:

(1)

where , , and capture the trade-off between the constraints[4].

As discussed previously, the basic multi-modality manifold-ranking model lacks sensitivity to the context in which the words appear since it is solely based on the BOW assumption. It ignores the internal structure of the sentences and does not consider word orders. Our aim in this paper is to propose a similarity measure in which syntactic and/or semantic information can be added to enhance the multi-modality manifold-ranking model by encoding the relational information between the words in sentences. We claim that for a complex task like topic-focused multi-document summarization where the relatedness between the document sentences is an important factor, the multi-modality manifold algorithm for ranking sentences would perform more effectively if we could incorporate the syntactic and semantic information with the standard cosine measure (i.e. TF*IDF) in calculating the similarity between sentences. In the next section, we describe how we can encode syntactic and semantic structures in calculating the similarity between sentences.

4 Syntactic and Shallow Semantic Structures

Given a sentence (or query[5]), we first parse it into a syntactic tree using a parser like (Charniak, 1999) and then, calculate the similarity between the two trees using the *tree kernel* (discussed in Section 4.1). However, syntactic information is often not adequate when dealing with long and articulated sentences or paragraphs. Shallow semantic representations, bearing a more compact information, could prevent the sparseness of deep structural approaches (Moschitti et al., 2007). Initiatives such as PropBank (PB) (Kingsbury and Palmer, 2002) have made possible the design of accurate automatic Semantic Role Labeling (SRL) systems like ASSERT (Hacioglu et al., 2003).

[3]We augment syntactic and/or semantic information with this measure in our proposed model using the syntactic and/or shallow semantic kernels described in Section 4 and argue that the combined measure performs better.

[4]The first two terms of the right-hand side in the cost function are the smoothness constraints for the two modalities while the last term denotes the fitting constraint.

[5]The query is denoted as the first point in the data space of the manifold ranking framework and represented by .

Figure 1: Example of semantic trees

For example, consider the PB annotation:

```
[ARG0 all][TARGET use][ARG1
the french franc][ARG2
as their currency]
```

Such annotation can be used to design a shallow semantic representation that can be matched against other semantically similar sentences, e.g.

```
[ARG0 the Vatican][TARGET use]
[ARG1 the Italian lira][ARG2
as their currency]
```

In order to calculate the semantic similarity between the sentences, we first represent the annotated sentence (or query) using the tree structures like Figure 1 which we call Semantic Tree (ST). In the semantic tree, arguments are replaced with the most important word-often referred to as the semantic head. We look for noun first, then verb, then adjective, then adverb to find the semantic head in the argument. If none of these is present, we take the first word of the argument as the semantic head. This reduces the data sparseness with respect to a typical cosine measure representation used in the basic multi-modality manifold-ranking model.

4.1 Tree Kernels

Once we build the trees (syntactic or semantic), our next task is to measure the similarity between the trees. For this, every tree is represented by an dimensional vector , where the i-th element is the number of occurrences of the i-th tree fragment in tree . The tree fragments of a tree are all of its sub-trees which include at least one production with the restriction that no production

Figure 2: (a) An example tree (b) The sub-trees of the NP covering "the press".

rules can be broken into incomplete parts. Figure 2 shows an example tree and a portion of its subtrees.

Implicitly we enumerate all the possible tree fragments . These fragments are the axis of this m-dimensional space. Note that this needs to be done only implicitly, since the number is extremely large. Because of this, (Collins and Duffy, 2001) defines the tree kernel algorithm whose computational complexity does not depend on .

The tree kernel of two trees and is actually the inner product of and :

(2)

We define the indicator function to be 1 if the sub-tree is seen rooted at node and 0 otherwise. It follows:

where, and are the set of nodes in and respectively. So, we can derive:

(3)

where, we define = .
Next, we note that can be computed in polynomial time, due to the following recursive definition:

1. If the productions at and are different then

2. If the productions at and are the same, and and are pre-terminals, then

3. Else if the productions at and are not pre-terminals,

(4)

where, is the number of children of in the tree; because the productions at and are the same, we have . The i-th child-node of is . *TK* is the similarity value (tree kernel) between the sentences (and/or the query sentence) based on the syntactic structure. For example, for the following sentence and query we get the following score:

Query (q): Describe steps taken and worldwide reaction prior to introduction of the Euro on January 1, 1999. Include predictions and expectations reported in the press.

Sentence (s): Europe's new currency, the euro, will rival the U.S. dollar as an international currency over the long term, Der Spiegel magazine reported Sunday.

Score: 65.5

4.2 Shallow Semantic Tree Kernel (SSTK)

The tree kernel (TK) function computes the number of common subtrees between two trees. Such subtrees are subject to the constraint that their nodes are taken with all or none of the children they have in the original tree. Though, this definition of subtrees makes the TK function appropriate for syntactic trees but at the same time makes it not well suited for the semantic trees (ST). The critical aspect of steps (1), (2) and (3) of the TK function is that the productions of two evaluated nodes have to be identical to allow the match of further descendants. This means that common substructures cannot be composed by a node with only some of its children as an effective ST representation would require. (Moschitti et al., 2007) solve this problem by designing the Shallow Semantic Tree Kernel (SSTK) which allows to match portions of a ST. The SSTK is based on two ideas: first, it changes the ST by adding *SLOT* nodes. These accommodate argument labels in a specific order i.e. it provides a fixed number of slots, possibly filled with *null* arguments, that encode all possible predicate arguments. Leaf nodes are filled with the wildcard character * but they may alternatively accommodate additional information. The slot nodes are used in such a way that the adopted TK function can generate fragments containing one or more children. As previously pointed out, if the arguments were directly attached to the root node, the kernel function would only generate the structure with all children (or the structure with no children, i.e. empty). Second, as the original tree kernel would generate many matches with slots filled with the null label, we have set a new step 0 in the TK calculation:
(0) if (or) is a pre-terminal node and its child label is *null*, ;
and subtract one unit to , in step 3:

The above changes generate a new *C* which, when substituted (in place of original *C*) in Eq. 3, gives the new SSTK. For example, for the following sentence and query we get the semantic score:

Query (q): Describe steps taken and worldwide reaction prior to introduction of the Euro on January 1, 1999. Include predictions and expectations reported in the press.

Sentence (s): The Frankfurt-based body said in its annual report released today that it has decided on two themes for the new currency

history of European civilization and abstract or concrete paintings.

Score: 9

5 Experiments and Results

5.1 Task Description

In this paper, we re-implement the multi-modality manifold ranking algorithm for topic-focused multi-document summarization by encoding the syntactic and semantic information to measure sentence relationships. We use the linear approach for fusing the modalities as this was shown to perform the best (Wan and Xiao, 2009). The purpose of our experiments is to study the impact of the syntactic and semantic representation in the multi-modality manifold-ranking framework.

Over the past three years, complex questions have been the focus of much attention in both the automatic question-answering and multi-document summarization (MDS) communities. While most current complex QA evaluations (including the 2004 AQUAINT Relationship QA Pilot, the 2005 Text Retrieval Conference (TREC) Relationship QA Task, and the 2006 GALE Distillation Effort) require systems to return unstructured lists of candidate answers in response to a complex question, recent MDS evaluations (including the 2005, 2006 and 2007 Document Understanding Conferences (DUC)) have tasked systems with returning paragraph-length answers to complex questions that are responsive, relevant, and coherent. The DUC conference series is run by the National Institute of Standards and Technology (NIST) to further progress in summarization and enable researchers to participate in large-scale experiments. We use the main task of DUC 2006 for evaluation. The task was: "Given a complex question (topic description) and a collection of relevant documents, the task is to synthesize a fluent, well-organized 250-word summary of the documents that answers the question(s) in the topic". To accomplish this task, we generate summaries for a subset of 10 topics of DUC 2006 dataset by each of our six systems as defined below:

(1) COSINE: This system is the original multi-modality manifold ranking method described in Section 3 that uses the standard cosine similarity measure based on TF*IDF and does not consider the syntactic/semantic information.

(2) SYN: This system measures the similarity between the sentences using the *syntactic tree* and the *general tree kernel* function defined in Section 4.1.

(3) SEM: This system measures the similarity between the sentences using the *shallow semantic tree* and the *shallow semantic tree kernel* function defined in Section 4.2.

(4) COSINE+SYN: This system measures the similarity between the sentences using both standard cosine similarity measure and the syntactic tree kernel.

(5) COSINE+SEM: This system measures the similarity between the sentences using both standard cosine similarity measure and the shallow semantic tree kernel.

(6) COSINE+SYN+SEM: This system measures the similarity between the sentences using standard cosine similarity measure, syntactic tree kernel, and shallow semantic tree kernel.

5.2 Automatic Evaluation

We carried out automatic evaluation of our candidate summaries using ROUGE (Lin, 2004) toolkit, which has been widely adopted for automatic summarization evaluation. *ROUGE* stands for "Recall-Oriented Understudy for Gisting Evaluation". It is a collection of measures that determines the quality of a summary by comparing it to reference summaries created by humans. The measures count the number of overlapping units such as n-gram, word-sequences, and word-pairs between the system-generated summary to be evaluated and the ideal summaries created by humans. For all our systems, we report the widely accepted important metrics: ROUGE-2 and ROUGE-SU. We also present the ROUGE-1 scores since this has a high correlation with the human judgement. All the ROUGE measures were calculated by running ROUGE-1.5.5 with stemming but no removal of stopwords. ROUGE run-time parameters were set as the same as DUC 2007 evaluation setup. They are:

ROUGE-1.5.5.pl -2 -1 -u -r 1000 -t 0 -n 4 -w 1.2 -m -l 250 -a

Table 1 to Table 3 show the ROUGE-1, ROUGE-2, and ROUGE-SU scores of our six different systems. In the experiments, the regularized parameter for the fitting constraint is fixed at 0.4, as in Wan et al. (2007). We kept as it was shown to be the most effective choice for the

linear fusion scheme in Wan and Xiao (2009).

Systems	Recall	Precision	F-score
COSINE	0.3619	0.3043	0.3305
SYN	0.3571	0.3105	0.3320
SEM	0.3814	0.2909	0.3299
COSINE+SYN	0.3627	0.3105	0.3346
COSINE+SEM	0.3737	0.3140	0.3412
COSINE+SYN+SEM	0.3648	0.3117	0.3360

Table 1: ROUGE-1 measures

Systems	Recall	Precision	F-score
COSINE	0.0584	0.0488	0.0532
SYN	0.0638	0.0558	0.0595
SEM	0.0732	0.0555	0.0631
COSINE+SYN	0.0611	0.0522	0.0563
COSINE+SEM	0.0691	0.0581	0.0631
COSINE+SYN+SEM	0.0658	0.0560	0.0605

Table 2: ROUGE-2 measures

Systems	Recall	Precision	F-score
COSINE	0.1262	0.0890	0.1043
SYN	0.1190	0.0903	0.1025
SEM	0.1406	0.0818	0.1033
COSINE+SYN	0.1278	0.0937	0.1081
COSINE+SEM	0.1334	0.0944	0.1104
COSINE+SYN+SEM	0.1282	0.0939	0.1083

Table 3: ROUGE-SU measures

For all the systems, Table 4 shows the F-scores of the reported ROUGE measures. From these results, we clearly see the positive impact of syntactic and semantic information in the multi-modality manifold ranking method for topic-focused multi-document summarization. The SYN system improves the ROUGE-1 and ROUGE-2 scores over the COSINE system by 0.45%, and 11.84% while underperforms the ROUGE-SU score by 1.75% respectively. The SEM system improves the ROUGE-2 scores over the COSINE system by 18.60% while underperforms the ROUGE-1 and ROUGE-SU scores by 0.18%, and 0.96% respectively. The COSINE+SYN system improves the ROUGE-1, ROUGE-2, and ROUGE-SU scores over the COSINE system by 1.24%, 5.82%, and 3.64% respectively. The COSINE+SEM system improves the ROUGE-1, ROUGE-2, and ROUGE-SU scores over the COSINE system by 3.23%, 18.60%, and 5.84% respectively. Lastly, the COSINE+SYN+SEM system improves the ROUGE-1, ROUGE-2, and ROUGE-SU scores over the COSINE system by 1.66%, 13.72%, and 3.83% respectively. Deep analysis of all these results yields that the proposed systems (that encode the syntactic and/or semantic information in the multi-modality manifold ranking framework) considerablely outperform the standard cosine similarity based manifold approach. The results also denote that encoding the syntactic and/or semantic information on top of the standard cosine similarity measure often outperform the systems that consider only syntactic and/or semantic information. From all our six systems, we can see that the *SEM* and *COSINE+SEM* are the best performing systems on average while performance of the *COSINE+SYN+SEM* decreases a bit indicating the fact that encoding both syntactic and semantic information on top of the standard cosine similarity measure has a negative impact on the multi-modality manifold ranking method. This may be due to the fact that the SYN system does not perform too well as seen from the results and thus deteriorates the performance of the COSINE+SYN+SEM system.

Systems	R-1	R-2	R-SU
COSINE	0.3305	0.0532	0.1043
SYN	0.3320	0.0595	0.1025
SEM	0.3299	0.0631	0.1033
COSINE+SYN	0.3346	0.0563	0.1081
COSINE+SEM	0.3412	0.0631	0.1104
COSINE+SYN+SEM	0.3360	0.0605	0.1083

Table 4: ROUGE F-scores for different systems

In Table 5, the proposed methods are compared with the NIST baseline. The NIST baseline is the official baseline system established by NIST that generated the summaries by returning all the leading sentences (up to 250 words) in the
field of the most recent document(s). We also list the average ROUGE scores of all the participating systems for DUC-2006 (i.e. AverageDUC). From the tables, we can see that the proposed multi-modality manifold ranking methods based on the syntactic and semantic measures mostly outperform the NIST baseline system. They can also achieve higher ROUGE scores as comparable to the average scores of all the participating systems of DUC-2006.

Confidence Intervals We also show 95% confidence interval of the important evaluation metrics for our systems to report significance for doing meaningful comparison. We use the ROUGE tool for this purpose. ROUGE uses a randomized method named bootstrap resampling to com-

Systems	ROUGE-1	ROUGE-2
COSINE	0.3305	0.0532
SYN	0.3320	0.0595
SEM	0.3299	0.0631
COSINE+SYN	0.3346	0.0563
COSINE+SEM	0.3412	0.0631
COSINE+SYN+SEM	0.3360	0.0605
Baseline	0.3209	0.0526
AverageDUC	0.3778	0.0748

Table 5: System comparison (F-scores)

pute the confidence interval. Bootstrap resampling has a long tradition in the field of statistics (Efron and Tibshirani, 1994). We use 1000 sampling points in the bootstrap resampling. Table 6 reports the 95% confidence intervals of the important ROUGE measures.

Systems	R-2	R-SU
COSINE	0.0401 - 0.0682	0.0854 - 0.1207
SYN	0.0439 - 0.0802	0.0845 - 0.1313
SEM	0.0530 - 0.0753	0.0928 - 0.1128
COSINE+SYN	0.0366 - 0.0805	0.0918 - 0.1286
COSINE+SEM	0.0499 - 0.0799	0.0873 - 0.1328
COSINE+SYN+SEM	0.0436 - 0.0795	0.0949 - 0.1205

Table 6: 95% confidence intervals for different systems

5.3 Manual Evaluation

Even if the ROUGE scores had significant improvement, it is possible to make bad summaries that get state-of-the-art ROUGE scores (Sjöbergh, 2007). So, we conduct an extensive manual evaluation in order to analyze the effectiveness of our systems. Two university graduate students judged the summaries for linguistic quality and overall responsiveness according to the DUC-2007 evaluation guidelines[6]. The given score is an integer between 1 (very poor) and 5 (very good) and is guided by consideration of the following factors: 1. Grammaticality, 2. Non-redundancy, 3. Referential clarity, 4. Focus and 5. Structure and Coherence. They also assigned a content responsiveness score to each of the automatic summaries. The content score is an integer between 1 (very poor) and 5 (very good) and is based on the amount of information in the summary that helps to satisfy the information need expressed in the topic. Table 7 presents the average linguistic quality and overall responsive scores of all our systems. These results also justify our claim by showing positive impacts of encoding syntactic and/or semantic information in the multi-modality manifold ranking framework. From these results, we can see that the proposed syntactic and/or semantic measure based systems outperform the COSINE system by a considerable margin.

[6]http://www-nlpir.nist.gov/projects/duc/duc2007/quality-questions.txt

Systems	Lin. Quality	Responsiveness
COSINE	2.50	3.60
SYN	3.40	3.80
SEM	4.10	4.40
COSINE+SYN	3.50	4.00
COSINE+SEM	2.60	3.40
COSINE+SYN+SEM	4.00	4.30

Table 7: Linguistic quality and responsiveness scores

6 Conclusion

In this paper, we proposed to encode the syntactic and semantic information for measuring sentence relationships in the multi-modality manifold ranking algorithm for topic-focused multi-document summarization and reported that adding syntactic and/or semantic information on top of the standard cosine measure improves the performance over the cosine measure alone. We parsed the sentences into the syntactic trees using the Charniak parser and applied the general tree kernel functions to measure the similarity between sentences. We used the shallow semantic tree kernel to measure the semantic similarity between two semantic trees. To the best of our knowledge, no other study has used syntactic and semantic information in the multi-modality manifold ranking model to improve its performance. We evaluated our systems automatically using ROUGE and conducted an extensive manual evaluation. Experimental results proved our claim by showing the effectiveness of the proposed methods.

Acknowledgments

We thank the anonymous reviewers for their useful comments. The research reported in this paper was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada – discovery grant and the University of Lethbridge.

References

E. Charniak. 1999. A Maximum-Entropy-Inspired Parser. In *Technical Report CS-99-12*, Brown University, Computer Science Department.

M. Collins and N. Duffy. 2001. Convolution Kernels for Natural Language. In *Proceedings of Neural Information Processing Systems*, pages 625–632, Vancouver, Canada.

H. Daumé III and D. Marcu. 2006. Bayesian query-focused summarization. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 305–312.

B. Efron and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. CRC Press.

K. Hacioglu, S. Pradhan, W. Ward, J. H. Martin, and D. Jurafsky. 2003. Shallow Semantic Parsing Using Support Vector Machines. In *Technical Report TR-CSLR-2003-03*, University of Colorado.

T. Hirao, , J. Suzuki, H. Isozaki, and E. Maeda. 2004. Dependency-based Sentence Alignment for Multiple Document Summarization. In *Proceedings of COLING 2004*, pages 446–452, Geneva, Switzerland. COLING.

P. Kingsbury and M. Palmer. 2002. From Treebank to PropBank. In *Proceedings of the International Conference on Language Resources and Evaluation*, Las Palmas, Spain.

C. Y. Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of Association for Computational Linguistics*, pages 74–81, Barcelona, Spain.

I. Mani and M. Maybury, 1999. *Advances in Automatic Text Summarization*. MIT Press.

I. Mani, 2001. *Automatic Summarization*. John Benjamins Co, Amsterdam/Philadelphia.

A. Moschitti and R. Basili. 2006. A Tree Kernel Approach to Question and Answer Classification in Question Answering Systems. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genoa, Italy.

A. Moschitti, S. Quarteroni, R. Basili, and S. Manandhar. 2007. Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classificaion. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 776–783, Prague, Czech Republic. ACL.

V. Nastase. 2008. Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-08)*, pages 763–772.

V. Punyakanok, D. Roth, and W. Yih. 2004. Mapping Dependencies Trees: An Application to Question Answering. In *Proceedings of AI & Math*, Florida, USA.

J. Sjöbergh. 2007. Older Versions of the ROUGEeval Summarization Evaluation System Were Easier to Fool. *Information Processing and Management*, 43:1500–1505.

X. Wan and J. Xiao. 2009. Graph-based multi-modality learning for topic-focused multi-document summarization. In *Proceedings of the 21st international jont conference on Artifical intelligence (IJCAI-09)*, pages 1586–1591.

X. Wan, J. Yang, and J. Xiao. 2007. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of the 20th international joint conference on Artifical intelligence (IJCAI-07)*, pages 2903–2908.

A. Zhang and W. Lee. 2003a. Question Classification using Support Vector Machines. In *Proceedings of the Special Interest Group on Information Retrieval*, pages 26–32, Toronto, Canada. ACM.

D. Zhang and W. S. Lee. 2003b. A Language Modeling Approach to Passage Question Answering. In *Proceedings of the Twelfth Text REtreival Conference*, pages 489–495, Gaithersburg, Maryland.

D. Zhou, O. Bousquet, T. Navin Lal, J. Weston, and B. Schölkopf. 2003a. Learning with local and global consistency. In *Proceedings of NIPS-03*.

D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. 2003b. Ranking on data manifolds. In *Proceedings of NIPS-03*.