A Hybrid Approach to Utilize Rhetorical Relations for Blog Summarization

Shamima Mithun Leila Kosseim
Concordia University
Department of Computer Science and Software Engineering
Montreal, Quebec, Canada
{s_mithun, kosseim}@encs.concordia.ca

Abstract. The availability of huge amounts of online opinions has created a new need to develop effective query-based opinion summarizers to analyze this information in order to facilitate decision making at every level. To develop an effective opinion summarization approach, we have targeted to resolve specifically Question Irrelevancy and Discourse Incoherency problems which have been found to be the most frequently occurring problems for opinion summarization. To address these problems, we have introduced a hybrid approach by combining text schema and rhetorical relations to exploit intra-sentential rhetorical relations. To evaluate our approach, we have built a system called BlogSum and have compared BlogSum-generated summaries after applying rhetorical structuring to BlogSum-generated candidate sentences without utilizing rhetorical relations using the Text Analysis Conference (TAC) 2008 data for summary contents. Evaluation results show that our approach improves summary contents by reducing question irrelevant sentences.

Keywords: Blog Summarization, Rhetorical Relations, Text Schema.

1 Introduction

Nowadays, because of the rapid growth of the Social Web, a large amount of informal opinionated texts are available on every topic. Natural language tools for automatically analyzing these opinions become necessary to help individuals, organizations, and governments to make timely decisions. For example, businesses and organizations are interested to know consumers' opinions and sentiments as part of their product and service evaluations ; individuals are interested to know others' opinions when they are intended to purchase some products or services... Query-based opinion summarizers from opinionated documents, as introduced in 2008 at the Text Analysis Conference (TAC), can address this need. Query-based opinion summarizers present what people think or feel on a given topic in a condensed manner to analyze others' opinions regarding a specific question (e.g. *Why do people like Starbucks better than Dunkin Donuts ?*). This research interest motivated us to develop an effective query-based multi-document opinion summarization approach for blogs.

The TAC 2008 summarization results show that blog summarizers typically perform weaker than news summarizers (Mithun & Kosseim, 2009). To analyze this in greater detail, we first tried to identify and categorize problems which typically occur in opinion summarization through an error analysis of the current blog summarizers. The goal of our research is to develop a blog summarization approach that addresses these most frequently occurring problems. For this error analysis, we used summaries from participating systems of the TAC 2008. Our study (Mithun & Kosseim, 2009) shows that *Question Irrelevancy*, *Topic Irrelevancy*, *Discourse Incoherency*, and *Irrelevant Information* are the most frequently occurring problems for blog summarization and the first three problems occur more frequently in blog summarization compared to news summarization. Figure 1 shows a sample summary (taken from the TAC 2008 opinion summarization track) that contains *Question Irrelevancy* and *Discourse Incoherency*.

Topic: Carmax
Question: What motivated positive opinions of Carmax from car buyers ?
Summary: At Carmax, the price is the price and when you want a car you go get one. Arthur Smith, 36, has been living in a van outside the Carmax lot, 24 hours a day, for more than a month. Sometimes I wonder why all businesses can't be like Carmax. [...]

FIGURE 1 – Sample Summary

The second sentence of the sample summary is not relevant to the question ; it exhibits a *Question Irrelevancy* problem. Moreover, in the summary, sentences are not interlinked ; as a result, they create a *Discourse Incoherency* problem. In our work, we target to deal with *Question Irrelevancy* and *Discourse Incoherency* and we believe our content selection approach may also reduce *Topic Irrelevancy*.

A successful query-based multi-document summarization approach needs to perform two tasks, namely content selection and content organization. For content selection, it needs to identify the most important information to be conveyed which are highly relevant to the question and should be incorporated in the summary to fulfil the information need. For content organization, the system needs to decide in which order the text should be presented because ordering has a significant effect on readers' comprehension of the summary. The problems of *Question Irrelevancy* and *Discourse Incoherency*, present mostly in query-based opinion summarization, are actually the result of poor content selection and content organization. If the summary contents are selected properly and the selected contents are organized properly then *Question Irrelevancy* and *Discourse Incoherency* problems should be significantly reduced.

To handle *Question Irrelevancy* and *Discourse Incoherency*, we have utilized rhetorical relations because successful text planning approaches, which require both content selection and content organization, use underlying rhetorical relations of text. Rhetorical relations have also been found useful in natural language generation (McKeown, 1985) and in news summarization (Blair-Goldensohn & McKeown, 2006; Bosma, 2004). In these news summarization work, inter-sentential rhetorical relations have been used. Due to the unavailability of automatic approaches to identify inter-sentential rhetorical relations, (Bosma, 2004)'s approach is domain dependent and (Blair-Goldensohn & McKeown, 2006) utilize only 2 types of rhetorical relations. However, to the best of our knowledge, rhetorical relations have never been used for blog summarization. In our work, to resolve *Question Irrelevancy* and *Discourse Incoherency* of blog summarization, we introduce a hybrid approach by combining text schema (McKeown, 1985), which is a standard text planning approach based on discourse organizing relations, with intra-sentential rhetorical relations. The following sections describe our approach and the evaluation results which show the effectiveness of our approach on blog summarization.

2 Blog Summarization based on Rhetorical Relations

To reduce *Question Irrelevancy* and *Discourse Incoherency* of blog summarization, we have adopted McKeown's text schema approach (McKeown, 1985). This approach is based on the observation that certain standard patterns of discourse organization (schema) are more effective to achieve a particular discourse goal (McKeown, 1985). For example, the definition of an object is often provided by a particular combination of sentence types ; whereas a comparison of two objects may use another combination to be effective. We also believe that for a particular type of question, certain types of sentences organized in a certain order can meet the communicative goal more effectively. As the text schema approach is designed to select relevant content and organize them coherently based on the underlying textual relations, we can make use of the schema-based framework.

A text schema shows that by using an organizational framework called schema (a combination of rhetorical predicates), a system can generate question relevant coherent multi-sentential texts given a communicative goal where rhetorical predicates characterize the structural purposes of texts and delineate the structural relations between propositions in a text. In a text schema-based approach, one can design and associate appropriate schemata (e.g. *compare and contrast*) to generate a summary that answers specific types of questions (e.g. *comparative*, *suggestion*). In the schema design, one can define constraints on the types of predicates (e.g. *analogy*, *condition*) and the order in which they should appear in the output summary for a particular question type. One can also specify constraints for each predicate of a schema to fulfill the communicative goal. For example, the Sample Schema of Table 1 specifies that the text should contain any number of comparative sentences followed by any number of contingency sentences. To be included in the summary, a sentence needs to be classified as either a *comparison* predicate or a *contingency* predicate. The sentence also needs to contain the specified argument (x), where the argument is the topic of the question, most often a named entity, and fulfil the specified constraints. The schema also specifies the order of sentences in the summary.

TABLE 1 – Sample Schema

Rhetorical Predicates	**Argument**	**Constraint**
Comparison*	(x)	the sentence compares (x) with anything else
Contingency*	(x)	the sentence has the same polarity as the question

We foresee that the schema should help to filter question irrelevant sentences by constraining what types of sentences can fill a particular slot of the schema and by imposing additional constraints for the sentence under a particular predicate type. In this approach, schemata should also help to improve coherency by removing question irrelevant sentences and also by specifying a higher level text organization by constraining on the order of the predicates.

The most challenging task in using a text schema-based approach for summarization is to identify which rhetorical predicate (e.g. *comparison*, *contingency*) is communicated by a candidate sentence in order to figure out if it should be included in the summary and where. In previous schema-based systems (e.g. (McKeown, 1985)), the application domain is typically represented as a knowledge base and the structure of the knowledge base is used to identify predicates. Predicates are also often identified by means of key words and other clues (e.g. *because*, *if*, *then*) or verb frames where a verb is associated with possible rhetorical predicates. To the best of our knowledge, there does not exist an approach to identify rhetorical predicates which is domain and genre independent.

Rhetorical predicates characterize the structural purpose of a proposition (e.g. the *attributive* predicate can describe the attribute of an object) or express the relationships that unite propositions (e.g. the *evidence* predicate creates a relation with the stated fact in order to provide support) (McKeown, 1985). We can see that rhetorical predicates can describe a single proposition or the relation between propositions. As rhetorical predicates and rhetorical relations described in various theories are comparable, we introduced a new way of identifying them for any domain. To identify rhetorical predicates, which characterize the relation between propositions - e.g. *evidence*, we used intra-sentential rhetorical relations with the help of the discourse parser SPADE (Soricut & Marcu, 2003). The SPADE parser is developed in the framework of RST (Mann & Thompson, 1988) and can automatically identify rhetorical relations within a sentence. In order to identify other types of rhetorical predicates, which characterize a single proposition on its own - e.g. *attributive*, we used a comparative classifier (Jindal & Liu, 2006) and the dependency relations of words (de Marneffe & Manning, 2008). We have built a system called BlogSum to test our approach.

3 BlogSum

Given an initial question on a particular topic and a set of related blogs, BlogSum performs two main tasks : content selection and content organization. The emphasis of our work is on content organization ; however, let us briefly discuss content selection before.

3.1 Content Selection

For content selection, BlogSum performs mainly sentence ranking to generate a ranked list of candidate sentences to be included in the summary. To rank sentences, BlogSum calculates a score for each sentence using the following features :

Sentence Score = Question Similarity + Topic Similarity + Subjectivity Score.

To calculate the question similarity, we used the cosine similarity between the sentence and the question. Sentences and questions are represented as a weighted word vector based on *tf.idf* (for sentences) and tf (for questions). The similarly between a sentence and the topic is calculated as with the question similarity using the words in the topic instead of the question words.

BlogSum uses the MPQA subjectivity lexicon [1], which contains more than 8000 entries of polarity words, to find the polarity of a word. In the lexicon, for each subjective word, the prior polarity and subjectivity strength (*weak*, *strong*) of the word are provided. In the lexicon, 4 types of prior polarity values are used namely, *positive*, *negative*, *both*, *neutral*. To assign a polarity to a word in a sentence, we used the polarity value *positive*, *negative*, or *both* and assign the score 1, -1, and 0.25, respectively. Moreover, if a word is tagged as weakly subjective then we reduce the subjectivity strength by 0.25, on the other hand, if a word is tagged as strongly subjective then we increase the subjectivity strength by 0.25. The subjectivity score of a sentence is then calculated based on the match of the sentence words with the subjective words listed in the subjectivity lexicon. Currently, the subjectivity score of a sentence is simply calculated in the following manner :

$$Subjectivity\ score\ of\ a\ sentence = \frac{sum\ of\ the\ polarity\ score\ of\ all\ sujective\ words\ found\ in\ the\ sentence}{\#\ of\ subjective\ words\ in\ the\ sentence}$$

Four types of polarity values including *positive*, *negative*, *mixed*, and *neutral* are used to classify a sentence. The subjectivity score of a sentence is used to determined its polarity class.

Positive	if	subjectivity score >= 0.5	**Mixed**	if	subjectivity score <0.5 to >-0.5
Negative	if	subjectivity score <= -0.5	**Neutral**	if	subjectivity score = 0 & no subjective word

The polarity of a sentence is identified and this information is used during the summary sentence selection. In general, the polarity of a sentence needs to be matched with the polarity of the question to be considered as a candidate summary sentence. For example, if the polarity of a question is positive then the polarity of a sentence also needs to be positive to become a summary sentence. The polarity of a question is calculated in the same way as of the polarity of a sentence. However, the subjectivity score of a question is only used

1. available at http ://www.cs.pitt.edu/mpqa/

to identify its polarity class but the subjectivity score of a sentence is used to identify its polarity class as well as calculating its rank. The content selection approach should also help to reduce *Topic Irrelevancy* as “topic similarity” is used as a ranking feature for sentence scoring.

3.2 Content Organization

The role of content organization is to select a few sentences from the candidate sentences and order them so as to produce a coherent and query relevant summary. For the purpose of content organization using the text schema-based approach, BlogSum performs the following main tasks :

1. Question Categorization,
2. Schema Selection, and
3. Predicate Identification.

In these tasks, questions need to be categorized based on their communicative goals and the most appropriate schema needs to be selected for the question categories. To incorporate candidate sentences in the final summary, according to the matched schema, sentences need to be classified into predefined rhetorical predicates. Let us now explain these tasks in more detail.

3.2.1 Question Categorization

In the schema-based approach, the question categorization process is key, as each question type determines which schema will better convey the expected communicative goal of the answer and should be used for text planning. By analyzing the TAC 2008 opinion summarization track questions manually, we have categorized them into 3 categories based on their communicative goals namely *comparative*, *suggestion*, and *reason*. *Comparative* questions request about the difference between objects ; *suggestion* questions request for suggestions to solve some problems ; and *reason* questions request for reasons for some claims. Examples for *comparative*, *suggestion*, and *reason* type questions are given below :

1. Comparative - e.g. *Why do people like Starbucks better than Dunkin Donuts ?*
2. Suggestion - e.g. *What do Canadian political parties want to happen regarding NAFTA ?*
3. Reason - e.g. *Why do people like Mythbusters ?*

3.2.2 Schema Selection

We have designed three schemata, one for each question type, 1) *comparative*, 2) *suggestion*, and 3) *reason*. To design these schemata, we have analyzed 50 summaries generated by participating systems at the TAC 2008 opinion summarization track. From our analysis, we have derived which question types should contain which type of predicates. Each schema is designed based on giving priority to its associated question type and subjective sentences as we are generating summaries for opinionated texts. For each type of schema, we have also defined appropriate constraints for its predicates where these constraints are identified from our summary analysis. These schemata, as defined, are certainly not the only way to be designed ; however, they offer enough flexibility to generate different summaries given different strategies to select candidate sentences.

3.2.3 Predicate Identification

To fill in the selected schema for a particular question type using candidate sentences to generate the output summary, each sentence needs to be classified into a predefined set of rhetorical predicates ; we called this process, predicate identification. For predicate identification, we first defined a set of rhetorical predicates to be used ; then we devised an approach to classify candidate sentences into these rhetorical predicates.

Rhetorical Predicates
Five main types of rhetorical predicates were considered :

1. Attributive : Provides details about an entity or event. It can be used to illustrate a particular feature about a concept - e.g. *Mary has a pink coat.*
2. Comparison : Gives a comparison and contrast among different situations - e.g. *Perhaps that's why for my European taste Starbucks makes great espresso while Dunkin's stinks.*
3. Contingency : Provides cause, condition, reason, evidence for a situation, result or claim - e.g. *The meat is good because they slice it right in front of you.*
4. Illustration : Is used to provide additional information or detail about a situation - e.g. *Allied Capital is a closed-end management investment company that will operate as a business development concern.*
5. Attribution : Can characterize the rhetorical relation *attribution* where instances of reported speech both direct and indirect are used to mark the *attribution* relation. This predicate can also be used to express feelings, thoughts, or hopes - e.g. *I said actually I think Zillow is great.*

Three of these predicates also subsume other predicates as shown below :

Comparison : *Contrast*, *analogy*, and *preference.*

Illustration : *Joint*, *list*, *disjoint*, and *elaboration.*

Contingency : *Explanation*, *evidence*, *reason*, *cause*, *result*, *consequence*, *background*, *circumstance*, *condition*, *hypothetical*, *enablement*, and *purpose*.

Rhetorical relations characterized by *comparison*, *illustration*, and *contingency* predicates are also considered by the PDTB research group (Prasad *et al.*, 2008) and by (Carlson & Marcu, 2001). We consider two additional classes of predicates *attributive* and *attribution*. The *attributive* predicate, also included in Grimes' predicates (McKeown, 1985), is considered because of its capability of describing attributes or features of an object or event which is used quite often to answer the types of questions we are dealing with. The rhetorical relation modelled by the *attribution* predicate, also listed in (Carlson & Marcu, 2001), was considered because it is often used to capture the discourse relations in opinionated texts as it can be used to include feelings and thoughts (Carlson & Marcu, 2001). In building our predicate model, we considered all main rhetorical relations listed in Mann and Thompson's RST taxonomy (Mann & Thompson, 1988). These predicates are also considered in Grimes' and Williams' predicate lists (McKeown, 1985).

Sentence Tagging
Once we have defined our inventory of predicates, candidate sentences now need to be classified into these predicates. To identify rhetorical predicates which describe relations between propositions, we have used the discourse parser SPADE (Soricut & Marcu, 2003). In order to identify other types of rhetorical predicates, which describe a proposition on its own, we have used a comparative classifier (Jindal & Liu, 2006) and the dependency relations of words (de Marneffe & Manning, 2008).

The discourse parser SPADE (Soricut & Marcu, 2003) was developed in the framework of RST. In SPADE, a large number of fine grained rhetorical relations are considered compared to those in RST. The SPADE parser identifies discourse relations within a sentence by first identifying elementary discourse units (EDU)s, then identifying rhetorical relations between two EDUs (clauses) by following the RST theory. For example, the SPADE parser identifies two EDUs *[perhaps that's why for my european taste Starbucks makes great espresso] [while Dunkin's stinks]* for the sentence "perhaps that's why for my European taste Starbucks makes great espresso while Dunkin's stinks" and assigns the relation *contrast* between these two EDUs. In this process, each sentence processed by the SPADE parser will be labelled with its rhetorical relations. BlogSum uses these relations to classify a sentence into the corresponding rhetorical predicate. This may result in tagging a sentence with no or with multiple rhetorical predicates. In summary generation, the sentence can be selected by the schema based on any of the matched predicate it contains.

The SPADE parser can only identify predicates across text spans, and cannot identify those occurring within a single span. For example, in "*Dunkin Donuts' coffee tasted better than Starbucks*" a *comparison* predicate is used, but would not be identified by SPADE. However, in our analysis, we found that comparisons do occur within a single text span. In order to classify a sentence as a comparison predicate within a single text span, we adapted Jindal et al.'s approach (Jindal & Liu, 2006). They proposed a supervised learning approach to identify comparative sentences. Using a set of keywords and annotated texts, their approach generates patterns for comparison sentence mining. Later these patterns are extended using class sequential rule mining (Jindal & Liu, 2006) and these extended patterns are used as features for a Naïve Bayes classifier. We have used their annotated dataset to build a similar comparative classifier for the identification of intra-sentence comparative predicates.

To identify the *attributive* predicates, which typically occur within a single text span, we have devised a set of heuristic rules by analyzing datasets containing attributive sentences (summary sentences from TAC-2008) such as topic terms need to be a subject or object of the verb. Dependency relations (de Marneffe & Manning, 2008) of words from the Stanford parser is used in this process. For example, in the sentence "Picasa displays the zoom percentage" the topic "picasa" is the subject ; there will be a dependency relation "nsubj" between "picasa" and the verb "displays".

To tag a sentence, we run the SPADE parser, Jindal et al.'s approach and the dependency relations simultaneously. In this process, a sentence can be tagged by more than one approach and receive multiple tags. With an analysis of 221 random summary sentences from the TAC 2008 opinion summarization track, we have found that 71%, 6%, and 18% of the sentences were tagged by the SPADE parser, Jindal et al.'s approach, and the dependency relations, respectively. In this analysis, we have also found that 5% of the sentences received no tag and 31% were tagged with multiple predicates.

As an example of how our approach works, let us turn back to the example of Figure 1 (Section 1). Once our approach has been applied, the question irrelevant sentence (2^{nd} sentence) in the summary will be filtered out because this sentence will not be identified as containing predicates prescribed by the *Reason* schema. The reason is that in general the sentence is in the attributive predicate form but it is not describing the topic (Carmax) which is a requirement to be considered as an attributive predicate. Hence, the 2^{nd} sentence will be excluded from the final summary even if its content score may be high.

4 Evaluation

To evaluate our summarization approach, BlogSum-generated summaries can be verified for content and linguistic qualities especially coherence and overall readability. The content evaluation would give an indication of the question relevance of the summary as well as the usefulness of our approach and the linguistic quality evaluation would give an indication of the coherency of the summary. To date, we have evaluated BlogSum-generated summaries for content evaluation only. As a baseline, we used the original ranked list of sentences (OList) before applying rhetorical relations and compared them to the final summaries after the rhetorical structuring. We have used the data from TAC 2008 opinion summarization track for the evaluation. In this experiment, we used the ROUGE metric, which is a standard automatic summary content evaluation metric, using answer nuggets (provided by TAC), which had been created to evaluate participants' summaries at TAC, as gold standard summaries. Precision, recall, and F-score are calculated for BlogSum and OList using ROUGE-2 and ROUGE-SU4 for 38 questions on 20 topics. The ROUGE-2 score is based on the overlap of word bi-grams between the automatically generated summaries and gold standard summaries (Dunlavy *et al.*, 2007). The ROUGE-SU4 score is also based on the overlap of bi-grams between summaries but allows a maximum gap of 4 tokens between the two tokens in a bi-gram (skip-bi-gram), and includes uni-gram co-occurrence statistics as well (Dunlavy *et al.*, 2007). In this experiment, ROUGE scores are also calculated for all 19 submissions in TAC-2008 opinion track which did not use answer-snippets (answer-snippets were extracted by the participating QA systems at TAC 2008 QA track) in summary generation as we did not use answer-snippets for summarization. The evaluation results are shown in Table 2. Note that in the table *R :#* refers to the rank of the system compared to the other 19 systems.

TABLE 2 – Evaluation Results

	ROUGE-2			**ROUGE-SU4**		
System Name	**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
BlogSum	0.045	0.210	0.070 *(R : 1)*	0.022	0.390	0.036 *(R : 6)*
OList	0.041	0.230	0.066 *(R : 3)*	0.017	0.440	0.028 *(R : 10)*
Best	0.047	0.156	0.069	0.045	0.226	0.062
Average	0.029	0.165	0.043	0.021	0.369	0.028

As Table 2 shows, BlogSum achieved a better F-Measure as well as a better precision for ROUGE-2 and ROUGE-SU4 compared to OList. On the other hand, recall has dropped by some acceptable range. This was to be expected because we are filtering sentences from the candidate list. From the results, we can see that using ROUGE-2 BlogSum gained 6% in F-Measure over OList, 10% in precision but dropped by 9% recall. For ROUGE-SU4, BlogSum gained 28% F-Measure over OList with 29% gain of precision and 11% recall drop. In both cases, the overall F-Measure scores in BlogSum seems to have improved when compared to OList. Further observations show that for ROUGE-SU4, the OList's precision is lower than the average precision score ; whereas BlogSum's score is above average. Compared to the other systems, BlogSum achieved very good scores for ROUGE-2 ; it outperformed the best system in TAC 2008 for F-Measure. BlogSum also achieved competitive results using ROUGE-SU4, it ranked 6 out of 19 systems. From the ROUGE-SU4 precision value and a manual analysis of BlogSum summaries on 5 topics, we found that BlogSum still contains about 43% of question irrelevant sentences. We need to investigate the reasons for the presence of these sentences ; it could be that incorrect results of other intermediate tasks such as predicate identification, polarity identification, schema result in these irrelevant sentences. However, the manual analysis also shows that BlogSum reduced 21% of the question irrelevant sentences from OList.

5 Related Work

Influenced by McKeown's pioneering work (McKeown, 1985) many researchers have used text schema for coherent text generation where they defined predicates and designed schema according to their applications. However, the schema-based approaches are typically domain dependent and the domain knowledge is explicitly represented in knowledge bases. Later on, predicates are identified and schemata are designed based on the hierarchical structure and relations in the knowledge base. As opposed to using the text schema approach for a particular domain with the help of a knowledge base, we have used a text schema-based approach in combination with rhetorical relations for any given domain in blog summarization. In the recent work, (Sauper & Barzilay, 2009) use schemata or templates to create texts for a given topic (e.g. American Film Actors) from the Internet (e.g. Wikipedia). Their approach uses human generated texts on that particular topic to create the schemata and to select content automatically. In contrast to generate topic-based summaries from structured documents (Wikipedia articles), our approach generates query-based summaries from unstructured documents (blogs). Moreover, in our work, we do not have any human generated summaries for learning.

Rhetorical relations of texts have been utilized for text planning in diverse domains to generate coherent texts as well as for text summarizations. Most notably (Marcu, 1997) used RST relations for single document summarization and proposed a discourse relation identification parsing algorithm. In some work (Blair-Goldensohn & McKeown, 2006; Bosma, 2004), rhetorical relations are exploited successfully for multi-document summarization. In these work, rhetorical relations across sentences are utilized. (Bosma, 2004) shows the effectiveness of RST relations to incorporate additional contextual information for the question. The evaluation was done on selected domains for which annotated RST relations were available. (Blair-Goldensohn & McKeown, 2006) used rhetorical relations for summarization successfully. However, due to the lack of availability of automatic rhetorical relations identification approaches, they only cover two rhetorical relations *cause* and *contrast*. Even though rhetorical relations across sentences are found useful for summarization, summarization approaches could not make use of most of these relations for open domain because of the unavailability of the automatic identification of these relations.

In news summarization, rhetorical relations across sentences were found useful and we tried to explore whether rhetorical relations are also useful for blog summarization, as blogs are different in content and structure compared to news. As automatic approaches to identify rhetorical relations across sentences are not available, in our work we only exploit rhetorical relations within a sentence. To exploit rhetorical relations for blog summarization, we have adopted a text schema-based approach. To overcome the domain dependency (knowledge base oriented development) of this approach, we introduced a hybrid approach which uses rhetorical relations within a sentence to classify sentences as part of the predicate identification of the schema-based summary generation.

6 Conclusion & Future Work

With the goal of developing an efficient opinion summarization approach, we targeted to resolve *Question Irrelevancy* and *Discourse Incoherency* which are the most frequently occurring problems for opinion summarization. To resolve these problems, we have exploited intra-sentential rhetorical relations and developed a combined approach using a text schema and rhetorical relations to overcome the domain dependency of text schema. We have used a combination of the SPADE parser along with a comparative classifier and dependency relations of words based on a dependency parser to identify rhetorical predicates. We have evaluated our approach using the TAC 2008 data using the ROUGE-2 and ROUGE-SU4 metrics and seems to obtain a gain in performance over the original candidate list of sentences using both measures.

From our evaluation, we conclude that our approach seems to have a positive effect on content selection. Moreover, as our approach improves the precision, results also demonstrate that it is reducing the *Question Irrelevancy* problems. We will be able to further evaluate the effectiveness of our approach for content selection as well as organization soon. In addition, we plan to evaluate the following intermediate tasks of BlogSum : 1) the predicate identification approach and in particular the contribution of each tagging strategy ; and 2) the schema that were developed. In the future, we also need to evaluate the linguistic quality of BlogSum-generated summaries to measure the coherency of the summary. Moreover, currently, we are using sentence ranking scores by giving priority to the higher rank sentences to determine the order among sentences of a particular predicate type (e.g. *comparison*). However, this ordering approach can lead to incoherent summaries. In the future, we have to devise a coherent sentence ordering approach which can ensure inter-sentence coherency.

7 Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments on an earlier version of the paper. This work was financially supported by the Natural Sciences and Engineering Council of Canada (NSERC).

Références

BLAIR-GOLDENSOHN S. & MCKEOWN K. (2006). Integrating Rhetorical-Semantic Relation Models for Query-Focused Summarization. In *Proceedings of the Document Understanding Conference (DUC) Workshop at NAACL-HLT 2006*, New York, USA.

BOSMA W. (2004). Query-Based Summarization using Rhetorical Structure Theory. In *15th Meeting of Computational Linguistics in the Netherlands CLIN*, p. 29–44, Leiden, Netherlands.

CARLSON L. & MARCU D. (2001). *Discourse Tagging Reference Manual*. Rapport interne ISI-TR-545, University of Southern California Information Sciences Institute.

DE MARNEFFE M. & MANNING C. (2008). The Stanford Typed Dependencies Representation. In *CrossParser '08 : Coling 2008 : Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, p. 1–8, Manchester, United Kingdom.

DUNLAVY D., O'LEARY D., CONROY J. & SCHLESINGER J. (2007). QCS : A System for Querying, Clustering and Summarizing Documents. *Information Processing Management*, **43**(6), 1588–1605.

JINDAL N. & LIU B. (2006). Identifying Comparative Sentences in Text Documents. In *SIGIR '06 : Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 244–251, Seattle, Washington, USA.

MANN W. & THOMPSON S. (1988). Rhetorical Structure Theory : Toward a Functional Theory of Text Organisation. *Text*, **3**(8), 234–281.

MARCU D. (1997). From Discourse Structures to Text Summaries. In *Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization*, p. 82–88, Madrid, Spain.

MCKEOWN K. (1985). Discourse Strategies for Generating Natural-Language Text. *Artificial Intelligence*, **27**(1), 1–41.

MITHUN S. & KOSSEIM L. (2009). Summarizing Blog Entries versus News Texts. In *Proceedings of Events in Emerging Text Types (eETTS). A Workshop of Recent Advances in Natural Language Processing RANLP 2009*, p. 35–42, Borovets, Bulgaria.

PRASAD R., MILTSAKAKI E., DINESH N., LEE A., JOSHI A., ROBALDO L. & WEBBER B. (2008). *The Penn Discourse Treebank 2.0. Annotation Manual*. Rapport interne IRCS-08-01, Institute for Research in Cognitive Science, University of Pennsylvania.

SAUPER C. & BARZILAY R. (2009). Automatically Generating Wikipedia Articles : A Structure-Aware Approach. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, p. 208–216, Suntec, Singapore.

SORICUT R. & MARCU D. (2003). Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *NAACL '03 : Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, p. 149–156, Edmonton, Canada.