SEMANTIC ROLE FRAMES GRAPH-BASED MULTIDOCUMENT SUMMARIZATION

Ercan Canhasi, Igor Kononenko
Laboratory for Cognitive Modeling
University of Ljubljana, Faculty of Computer and Information Science
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia

ABSTRACT

Multi-document summarization is a process of automatic creation of a compressed version of the given collection of documents. Recently, the graph-based models and ranking algorithms have been extensively researched by the extractive document summarization community. While most work to date focuses on sentence-level relations in this paper we present graph model that emphasizes not only sentence level relations but also the influence of under sentence level relations (e.g. a part of sentence similarity). By using the proven cognitive psychology model (the Event-Indexing model) and semantic role parsing for generating the frame graph, we establish the bases for distinguishing the sentence level relations. Based on this model, we developed an iterative frame and sentence ranking algorithm, based on the existing well known PageRank algorithm. Experiments are conducted on the DUC 2004 data sets and the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) evaluation results demonstrate the advantages of the proposed approach.

1 INTRODUCTION

Multi-document summarization (MDS) aims to filter the most important information from a set of documents to generate a compressed summary. Recently, the graph-based models and sentence ranking algorithms based on these models have been extensively researched.

Conventionally, they model a document or a group of documents as a text graph composed by taking a text unit, such as a term or a sentence as a node, and similarity or affinity between text units as edges. The significance of a node in a graph is then estimated by graph-based ranking algorithms which take into account global information recursively computed from the entire graph. Sentences in document(s) are ranked according to the computed node significance and the most important ones are selected to form an extractive summary.

While most work to date focuses on the sentence and the document level relations, in this work, considering importance of the intra sentence relations and being inspired by Event-Indexing model [10], a well known cognitive model for text understanding and representation, we present a new graph model, the frame graph. A frame graph is composed by taking semantic role frames [2] as nodes and similarities between frames as edges. The significance of a node in a frame graph is estimated by graph-based ranking algorithms.

The remainder of this paper is organized as follows. Section 2 reviews existing graph-based summarization models. Sections 3 and 4 introduce the proposed summarization method. After that, Section 5 reports experiments and evaluation results. Finally, Section 6 concludes the paper.

2 RELATED WORK

The graph-based models have drawn considerable attention from the extractive document summarization community in the past years [3,6].

So far, the most popular graph-based ranking algorithms applied in document summarization are PageRank [1] and HITS [4] and their variations. Those algorithms make use of "voting" or "recommendations" between sentences to evaluate the importance of sentences in the documents.

Erkan and Radev [3] represented documents as a weighted undirected graph by taking sentences as nodes and the cosine similarity between sentences as the edge weight function. An algorithm called LexRank, adapted from PageRank, was applied to calculate the sentence significance, which was then used as the criterion to rank and select summary sentences. Meanwhile, Mihalcea and Tarau [6] presented their PageRank variation, called TextRank, in the same year.

All above mentioned exemplary work was concerned with generic summarization. Later on, graph-based ranking algorithms were introduced in query-oriented summarization too, when this new challenge became a popular research topic recently. For example, a topic-sensitive version of PageRank was proposed in [7]. A variety of other graph-based methods have been proposed for topic-focused multi-document summarization [8,9]. Different from generic summarization, a query-oriented summarization is necessarily driven by queries.

3 MDS VIA SEMANTIC ROLE GRAPHS ARGUMENTS

The purpose of this study is to show that it is possible to improve the efficiency of summarization using a semantically richer representation. Here by richer representation we mean the semantic graph of a document or a set of documents. The document set $D = d_1, d_2, \dots, d_n$

is represented as a weighted undirected text graph G by taking parts of sentences (semantic frames) in D as vertices and adding an edge to connect the two vertices if the two frames of concerned sentences are similar enough.

3.1 Motivation for using semantic role graphs

Summarization task requires understanding the document and presenting the important parts. In extractive summarization, this task is achived by selecting the sentences to be included in the summary. The most common method to solve this problem is to rank the sentences according to their informativeness.

Since humans tend to include sentences containing most frequent words in their summaries, the word-based frequency calculations for sentence scoring are baseborn approachs for MDS. However, this approach is semantically incomplete, because words alone usually do not carry semantic information. On the other hand, even if humans do not always agree on the content to be added to a summary, they perform very well on this task. Therefore our goal should be to find a way of mimicking the cognition behind the human like summarization process.

Our motivation for using SRL frames in sentence scoring for MDS originates from given concerns. Instead of using individual terms for sentence scoring, we exploit semantic arguments and relations between them by using the psychology cognitive situation model, namely the Event-Indexing model.

3.2 Event Indexing Model and Semantic Role Labeler

According to Event-indexing model a human-like system should keep track of five indices while reading the document. Those indices are *protagonist*, *temporality*, *spatiality*, *causality* and *intention*, with the given descending order of importance. One can also show that the semantic role parser's output can be mapped to the above proposed cognitive model.

Semantic roles are defined as the relationships between syntactic constituents and the predicates. Most sentence components have semantic connections with the predicate, carrying answers to the questions such as who, what, when, where etc. From the aspect of the semantic parser, frame arguments can be mapped to cognitive model indices as follows:

- A protagonist can be found in an answer to question "who", or more precisely in arguments A0 or A1 or A2. Argument A0 is the subject of the frame, as shown in Table 1, A1 is the object and A2 is the indirect object. Although in original work the protagonist is defined as a person around whom the story takes place, we see it reasonable to expand the notion of protagonist to the subject or object that can be everything, from a person to an organization or some abstract concept.
- Temporality is the temporal information in each frame and can be extracted from the frame argument AM_{TMP}.
- Spatiality is the space or location information of each frame and is equal to argument AM_{LOC}.
- Causality indexing is concerned with actions of frames so it can be mapped to the frame predicate.
- The intentionality-indexing is quite vague but since its weight of significance is less than of the others, as defined in the original work, we decided to omit it in this early versions of the system.

The SRL parser takes each sentence in the document set and properly labels the semantic word phrases. We refer to these phrases as semantic arguments or shortly arguments. There is an issue related to the SRL parsing process that we should take into account. For each verb in a sentence, the SRL parser provides a different frame. It considers the verb as the predicate of the sentence and tries to label the remaining part of the sentence as proper arguments. However, if the selected verb is not the actual predicate, the parser fails to identify most of the words as a part of an argument. Therefore, we consider the frame that leaves the least number of terms unlabeled as the complete parse of the sentence. In our calculations we use also the rest of frames but we treat them as incomplete. Since we don't want to lose information that can be brought to the resulting graph, instead of eliminating partially parsed frames we use them, but with lower weight in the similarity calculation.

Arguments labeling		Arguments modifier	
rel	verb	AM_{ADV}	Adverb mod.
A0	Subject	AM_{DIR}	Direction
A1	Object	AM_{DIS}	Discourse mrk
A2	Ind. object	AM_{LOC}	Location
A3	Start point	AM_{MNR}	Manner
A4	End point	AM_{NEG}	Negation
A5	Direction	AM_{PRD}	Sec. Predicate
		AM_{PRP}	Purpose
		AM_{TMP}	Temporal mrk.

Table1: Representation of label arguments and modifiers.

4 PROPOSED METHOD

The summarization method, we propose, works in the following way, as illustrated in Fig.1. First, the documents are given to the SRL parser where the semantic arguments from each parsed sentence are extracted. We calculate the composite similarity between all semantic frames based on the event-indexing. Then we generate a semantic graph where nodes are semantic frames and edges are the composite similarity values.

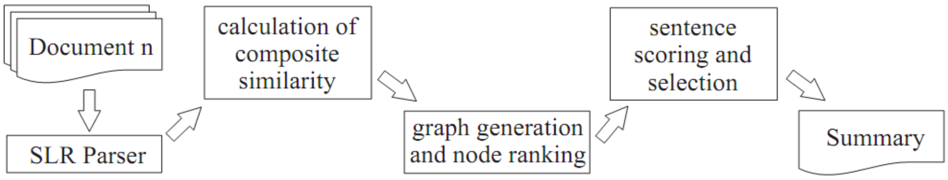

Figure 1: MDS System

Next we use the modified version of PageRank for identifying the significant edges in the graph. Later, we sum the PageRank scores of semantic frames, originating from the same sentence, and we use it as a score for sentence scoring. The next step aims to further remove redundant information in the summary by penalizing the sentences largely overlapping with other highly ranked sentences. Based on the text graph and the obtained rank scores, a greedy algorithm is applied to inflict the diversity penalty and compute the final rank scores of the sentences. Subsequently, the top scoring sentences are selected one-by-one and put into the summary.

4.1 Graph modeling

In this section we present our novel graph model, which is used in the frame ranking algorithm, presented in the next section. Let a set of documents D be a text similarity graph $G = (V, E, \alpha_v, \alpha_e)$ where V represents the frame vertex set, $E \subseteq V \times V$ is a frame edge set. $\alpha: V \longrightarrow \Re_+$ is defined to label frame vertices, while $\beta: E \longrightarrow \Re_+$ is a function for labeling frame edges.

Vertex Function	$\alpha(f_i) = \begin{cases} \frac{1}{N} \; if\ frame\ complete \\ \frac{1}{2N} \; otherwise \end{cases}$
Edge Function	$\beta(f_i, f_j) = sim_{composite}(f_i, f_j)$

Table 2: Different frames of a sentence.

The SRL frame edge function is formulated as the composite similarity function of two frames f_i and f_j. Let N be the total number of frames in a documents set. The frame vertex function assigns to frame vertices the value of $\frac{1}{N}$ or $\frac{1}{2N}$, depending on their completeness, where incomplete frames have lower weight. In Table 2, N is the total number of the frames in a document set, and the similarity between any two frames is defined by the composite similarity function, which will be detailed in the next paragraphs.

Our goal is to capture the similarity and redundancy between sentences, but at a lower structural and a higher semantic level. To accomplish this, we use the event-indexing model as the base for calculations of semantic similarity between frames of semantic role parser outputs, namely frames. According to this model we need to define the similarity measure for protagonist (prt), temporality (tmp), spatiality (spt), and causality(cst).

$$sim_{prt}(f_i, f_j) = \alpha_1 sim(A0_i, A0_j) + \alpha_2 sim(A1_i, A1_j) + \alpha_3 sim(A2_i, A2_j) + \alpha_4 sim(A0_i, A1_j) + \alpha_5 sim(A0_i, A2_j) + \alpha_6 sim(A1_i, A2_j)$$

$$sim_{tmp}(f_i, f_j) = sim\left(AM_{TMP_i}, AM_{TMP_j}\right)$$

$$sim_{spt}(f_i, f_j) = sim\left(AM_{LOC_i}, AM_{LOC_j}\right)$$

$$sim_{cst}(f_i, f_j) = sim(Predicate_i, Predicate_j)$$

In order to have the flexible weighting scheme we use coefficients $\alpha_1 = \alpha_2 = \alpha_3 = 0.25$; $\alpha_4 = \alpha_5 = 0.10$; $\alpha_6 = 0.5$;

The compose similarity is defined as:

$$sim_{cmp}(f_i, f_j) = [\beta_1 sim_{prt}(f_i, f_j) + \beta_2 sim_{tmp}(f_i, f_j) + \beta_3 sim_{spt}(f_i, f_j) + \beta_4 sim_{cst}(f_i, f_j)] / \#arguments$$

where $\beta_1 = 0.4$; β_2=0.3; $\beta_3 = 0.2$; $\beta_4 = 0.1$. The values for coefficients are chosen based on the cognitive model which gives emphasis in the decreasing order to the protagonist, temporality, spatiality and causality. We also normalize the composite similarity value with the number of arguments used in the calculation of similarity.

4.1 Frame graph-based ranking algorithm

In previous section the idea of frame similarity graph is presented. Based on it, in this section we present a modified iterative graph-based sentence ranking algorithm. Our algorithm is extended from those existing PageRank-like algorithms reported in the literature that calculate the graph only in the sentence level [3,6].

In the summary, PageRank method (in matrix notation), as described in the original paper [1], is

$$\pi^{(k+1)T} = \alpha \pi^{(k)T} \boldsymbol{H} + \left(\alpha \pi^{(k)T} a + 1 - \alpha\right) \boldsymbol{v}^T$$

where $\boldsymbol{H}$ is a very sparse, raw sub stochastic frame similarity matrix, α is a scaling parameter between 0 and 1, π^T is the stationary row vector of $\boldsymbol{H}$ called the PageRank vector, $\boldsymbol{v}^T$ is a complete dense, rank-one teleportation vector and a is a binary dangling node vector. In terms of the sentence ranking the matrix $\boldsymbol{H}$ is an adjacency matrix of frames, $\boldsymbol{v}^T$ is the preference vector of frames and the resulting π^T is the frame ranking vector.

5 PLEMINARY RESULTS

The DUC[1] 2004 data set from DUC was tested to analyze the efficiency of the proposed summarization method. The Task 2 in the DUC 2004 is to generate a short summary (665 bytes) of an input set of topic-related news articles. The total number of document groups is 50, with each group containing 10 articles on average.

For each group, four NIST assessors were asked to create a brief summary. The machine-generated summaries are evaluated using ROUGE [5], the automatic n-gram matching which measures performance based on the number of co-occurrences between machine-generated and ideal summaries in different word units. The 1-gram ROUGE score (a.k.a.ROUGE-1) has been found to correlate very well with human judgments at a confidence level of 95%, based on various statistical metrics.

Even though in this version of method we did not consider sentence positions or other summary quality improvement techniques such as sentence reduction, its overall performance is promising, see Table 3. The use of the frame graph model in summarization can make considerable improvements even though the results presented here do not report a significant difference.

System	ROUGE-1[2]
Avg. of human assessors	0.403[0.383,0.424]
Best Machine (SYSID=65)	0.382[0.369,0.395]
Median Machine (SYSID=138)	0.343[0.328,0.358]
Worst machine (SYSID=111)	0.242[0.230,0.253]
Our model	0.374[0.359,0.389]
LexRank	0.369[0.354,0.382]
2 (NIST Baseline) (Rank 25/35)	0.324[0.309,0.339]
Random baseline:	0.315[0.303,0.328]

Table 3: ROUGE-1 scores of the DUC 2004 and evaluation of our model.

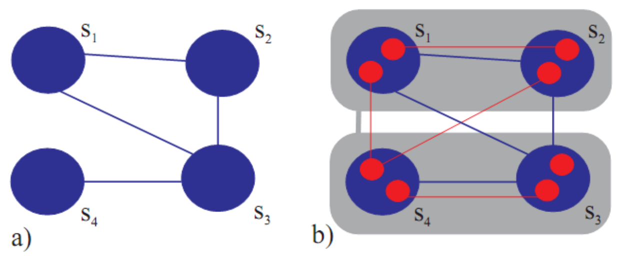

Figure 2: Summarization graph model (a) before and (b) after introducing multilayered model.

6 DISCUSSION AND FUTURE WORK

We have presented a frame graph model and a ranking algorithm for generic MDS. The main contribution of our work is introducing the concept of the frame graph model. The results of applying this model in extractive summarization are quite promising. There is work still left to be done, however. While most work to date focuses on homogeneous connectedness of sentences and heterogeneous connectedness of documents and sentences (e.g. sentence similarity weighted by document importance), in the future we hope to be able to present a novel 3-layered graph model that emphasizes not only sentence and document level relations but also the influence of under sentence level relations (e.g. a part of sentence similarity). By using an intelligent weighting scheme we plan to add two more layers, namely the sentence and the document layers, to frame graph model presented in this work (see Figure 2), which will yield to a richer multilayered graph model with the inter and intra sentence and the documental level relations. Currently, we are also working on further improvements of the model, and its adaptation to other summarization tasks, such as query and update summarizations.

References

[1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. Computer Networks, 30(1-7):107-117, 1998.

[2] Xavier Carreras and Lluis Marque. Introduction to the conll-2004 shared task: Semantic role labeling. In CoNLL, pages 89-97, 2004.

[3] Gunes Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. (JAIR), 22:457-479, 2004.

[4] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. J. ACM, 46(5):604-632, 1999.

[5] Chin-Yew Lin and Eduard H. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In HLT-NAACL, 2003.

[6] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In EMNLP, pages 404-411, 2004.

[7] Jahna Otterbacher, Gunes Erkan, and Dragomir R. Radev. Biased lexrank: Passage retrieval using random walks with question-based priors. Inf. Process. Manage., 45(1):42-54, 2009.

[8] Xiaojun Wan. Document-based hits model for multi-document summarization. In Tu Bao Ho and Zhi-Hua Zhou, editors, PRICAI, volume 5351 of Lecture Notes in Computer Science, pages 454-465. Springer, 2008.

[9] Furu Wei, Wenjie Li, Qin Lu, and Yanxiang He. A document-sensitive graph model for multi-document summarization. Knowl. Inf. Syst., 22(2):245-259, 2010.

[10] R.A. Zwaan, M.C. Langston, and A.C. Graesser. The construction of situation models in narrative comprehension: an event-indexing model.Psychological Science, 6(5) : 292-297, 1995.

[1] Document Understanding Conference(http://duc.nist.gov)
[2] 95% confidence interval