

Investigations on Event-Based Summarization

Mingli Wu

Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong
csmlwu@comp.polyu.edu.hk

Abstract

We investigate independent and relevant event-based extractive multi-document summarization approaches. In this paper, events are defined as event terms and associated event elements. With independent approach, we identify important contents by frequency of events. With relevant approach, we identify important contents by PageRank algorithm on the event map constructed from documents. Experimental results are encouraging.

1 Introduction

With the growing of online information, it is inefficient for a computer user to browse a great number of individual news documents. Automatic summarization is a powerful way to overcome such difficulty. However, the research literature demonstrates that machine summaries need to be improved further.

The previous research on text summarization can date back to (Luhn 1958) and (Edmundson 1969). In the following periods, some researchers focus on extraction-based summarization, as it is effective and simple. Others try to generate abstractions, but these works are highly domain-dependent or just preliminary investigations. Recently, query-based summarization has received much attention. However, it is highly related to information retrieval, another research subject. In this paper, we focus on generic summarization. News reports are crucial to our daily life. In this paper, we focus on effective summarization approaches for news reports.

Extractive summarization is widely investigated in the past. It extracts part of document(s) based on some weighting scheme, in which dif-

ferent features are exploited, such as position in document, term frequency, and key phrases. Recent extraction approaches may also employ machine learning approaches to decide which sentences or phrases should be extracted. They achieve preliminary success in different application and wait to be improved further.

Previous extractive approaches identify the important content mainly based on terms. Bag of words is not a good representation to specify an event. There are multiple possible explanations for the same collection of words. A predefined template is a better choice to represent the event. However it is domain-dependent and need much effort to create and fill it. This tension motivates us to seek a balance between effective implementation and deep understanding.

According to related works (Filatova and Hatzivassiloglou, 2004) (Vanderwende et al., 2004), we assume that event may be a natural unit to convey meanings of documents. In this paper, event is defined as the collection of event terms and associated event elements in clause level. Event terms express the meaning of actions themselves, such as “incorporate”. In addition to verbs, action nouns can also express meaning of actions and should be regarded as event terms. For example, “incorporation” is action noun. Event elements include named entities, such as person name, organization name, location, time. These named entities are tagged with GATE (Cunningham et al., 2002). Based on our event definition, independent and relevant event-based approaches are investigated in this research. Experiments show that both of them achieve encouraging results.

The related works are discussed in Section 2. Independent event-based summarization approach is described in Section 3. Relevant event-based summarization approach is described in Section 4. Section 5 presents the experiments and

evaluations. Then the strength and limitation of our approaches are discussed in Section 6. Finally, we conclude the work in Section 7.

2 Related Work

Term-based extractive summarization can date back to (Luhn, 1958) and (Edmundson, 1969). This approach is simple but rather applicable. It represents the content of documents mainly by bag of words. Luhn (1958) establishes a set of “significant” words, whose frequency is between a higher bound and a lower bound. Edmundson (1969) collects common words, cue words, title/heading words from documents. Weight scores of sentences are computed based on type/frequency of terms. Sentences with higher scores will be included in summaries. Later researchers adopt $tf*idf$ score to discriminate words (Brandow et al., 1995) (Radev et al., 2004). Other surface features are also exploited to extract important sentence, such as position of sentence and length of sentence (Teufel and Moens, 1999) (Radev et al., 2004). To make the extraction model suitable for documents in different domains, recently machine learning approaches are widely employed (Kupiec et al., 1995) (Conroy and Schlesinger, 2004).

To represent deep meaning of documents, other researchers have investigated different structures. Barzilay and Elhadad (1997) segment the original text and construct lexical chains. They employ strong chains to represent important parts of documents. Marcu (1997) describes a rhetorical parsing approach which takes unrestricted text as input and derives the rhetorical structure tree. They express documents with structure trees. Dejong (1978) adopts predefined templates to express documents. For each topic, the user predefines frames of expected information types, together with recognition criteria. However, these approaches just achieve moderate results.

Recently, event receives attention to represent documents. Filatovia and Hatzivassiloglou (2004) define event as action (verbs/action nouns) and named entities. After identifying actions and event entities, they adopt frequency weighting scheme to identify important sentence. Vanderwende et al. (2004) represent event by dependency triples. After analysis of triples they connect nodes (words or phrases) by way of semantic relationships. Yoshioka and Haraguchi (2004) adopt a similar approach to build a map, but they regard sentence as the nodes of the map.

After construction of a map representation for documents, Vanderwende et al. (2004), and Yoshioka and Haraguchi (2004) all employ PageRank algorithm to select the important sentences. Although these approaches employ event representation and PageRank algorithm, it should be noted that our event representation is different with theirs. Our event representation is based on named entities and event terms, without help of dependency parsing. These previous event-based approaches achieved promising results.

3 Independent Event-based Summarization

Based on our observation, we assume that events in the documents may have different importance. Important event terms will be repeated and always occur with more event elements, because reporters hope to state them clearly. At the same time, people may omit time or location of an important event after they describe the event previously. Therefore in our research, event terms occurs in different circumstances will be assigned different weights. Event terms occur between two event elements should be more important than event terms occurring just beside one event elements. Event terms co-occurring with participants may be more important than event terms just beside time or location.

The approach on independent event-based summarization involves following steps.

1. Given a cluster of documents, analyze each sentence one at a time. Ignore sentences that do not contain any event element.
2. Tag the event terms in the sentence, which is between two event elements or near an event element with the distance limitation. For example, [Event Element A, Event Term, Event Element B], [Event Term, Event Element A], [Event Element A, Event Term]
3. Assign different weights to different event terms, according to contexts of event terms. Different weight configurations are described in Section 5.2. Contexts refer to number of event elements beside event terms and types of these event elements.
4. Get the average $tf*idf$ score as the weight of every event term or event element. The algorithm is similar with Centroid.

5. Sum up the weights of event terms and event elements in a sentence.
6. Select the top sentences with highest weights, according to the length of summary.

4 Relevant Event-based Summarization

Independent event-based approaches do not exploit relevance between events. However, we think that it may be useful to identify important events. After a document is represented by events, relevant events are linked together. We made the assumption that important events may be mentioned often and events associated to important events may be important also. PageRank is a suitable algorithm to identify the importance of events from a map, according to the previous assumption. In the following sections, we will discuss how to represent documents by events and how to identify important event with PageRank algorithm.

4.1 Document Representation

We employ an event map to represent content of a document cluster, which is about a certain topic. In an event map, nodes are event terms or event elements, and edges represent association or modification between two nodes. Since the sentence is a natural unit to express meanings, we assume that all event terms in a sentence are all relevant and should be linked together. The links between every two nodes are undirectional.

In an ideal case, event elements should be linked to the associated event terms. At the same time, an event element may modify another element. For example, one element is a head noun and another one is the modifier. An event term (e.g., verb variants) may modify an event element or event term of another event. In this case, a full parser should be employed to get associations or modifications between different nodes in the map. Because the performance of current parsing technology is not perfect, an effective approach is to simulate the parse tree to avoid introducing errors of a parser. The simplifications are described as follows. Only event elements are attached with corresponding event terms. An event term will not be attached to an event element of another event. Also, an event element will not be attached to another event element. Heuristics are used to attach event elements with corresponding event terms.

Given a sentence “Andrew had become little more than a strong rainstorm early yesterday,

moving across Mississippi state and heading for the north-eastern US”, the event map is shown in Fig. 1. After each sentence is represented by a map, there will be multiple maps for a cluster of documents. If nodes from different maps are lexical match, they may denote same thing and should be linked. For example, if named entity “Andrew” occurred in Sentence A, B and C, then the three occurrences O_A , O_B and O_C will be linked as $O_A—O_B$, $O_B—O_C$, $O_C—O_A$. By this way, maps for sentences can be linked based on same concepts.

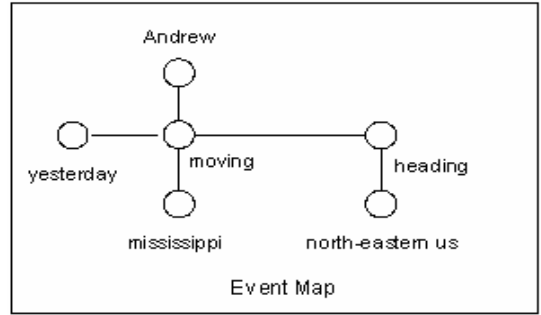


Figure 1. Document representation with event map

4.2 Importance Identification by PageRank

Given a whole map for a cluster of documents, the next step is to identify focus of these documents. Based on our assumption about important content in the previous section, PageRank algorithm (Page et al., 1998) is employed to fulfill this task. PageRank assumes that if a node is connected with more other nodes, it may be more likely to represent a salient concept. The nodes relevant to the significant nodes are closer to the salient concept than those not. The algorithm assigns the significance score to each node according to the number of nodes linking to it as well as the significance of the nodes. In PageRank algorithm, we use two directional links instead for every unidirectional link in Figure 1.

The equation to calculate the importance (indicated by PR) of a certain node A is shown as follows:

$$PR(A) = (1 - d) + d \left(\frac{PR(B_1)}{C(B_1)} + \frac{PR(B_2)}{C(B_2)} + \dots + \frac{PR(B_t)}{C(B_t)} \right)$$

Where B_1, B_2, \dots, B_t are all nodes which link to the node A . $C(B_i)$ is the number of outgoing links from the node B_i . The weight score of each node can be gotten by this equation recursively. d is the factor used to avoid the limitation of loop in the map structure. As the literature (Page et al., 1998) suggested, d is set as 0.85. The significance of each sentence to be included in the

summary is then derived from the significance of the event terms and event elements it contains.

5 Evaluation

5.1 Dataset and Evaluation Metrics

DUC 2001 dataset is employed to evaluate our summarization approaches. It contains 30 clusters and a total of 308 documents. The number of documents in each cluster is between 3 and 20. These documents are from some English news agencies, such as Wall Street Journal. The contents of each cluster are about some specific topic, such as the hurricane in Florida. For each cluster, there are 3 different model summaries, which are provided manually. These model summaries are created by NIST assessors for the DUC task of generic summarization. Manual summaries with 50 words, 100 words, 200 words and 400 words are provided.

Since manual evaluation is time-consuming and may be subjective, the typical evaluation package, ROUGE (Lin and Hovy, 2003), is employed to test the quality of summaries. ROUGE compares the machine-generated summaries with manually provided summaries, based on unigram overlap, bigram overlap, and overlap with long distance. It is a recall-based measure and requires that the length of the summaries be limited to allow meaningful comparison. ROUGE is not a comprehensive evaluation method and intends to provide a rough description about the performance of machine generated summary.

5.2 Experimental Configuration

In the following experiments for independent event-based summarization, we investigate the effectiveness of the approach. In addition, we attempt to test the importance of contextual information in scoring event terms. The number of associated event terms and the type of event terms are considered to set the weights of event terms. The weights parameters in the following experiments are chosen according to empirical estimations.

Experiment 1: Weight of any entity is 1. Weight of any verb/action noun, which is between two entities or just beside one entity, is 1.

Experiment 2: Weight of any entity is 1. Weight of any verb/action noun, which is between two entities, is 3. Weight of any verb/action noun, which is just beside one entity, is 1.

Experiment 3: Weight of any entity is 1. Weight of any verb/action noun, which is be-

tween two entities and the first entity is person or organization, is 5. Weight of any verb/action noun, which is between two entities and the first entity is not person and not organization, is 3. Weight of any verb/action noun, which is just after a person or organization, is 2. Weight of any verb/action noun, which is just before one entity, is 1. Weight of any verb/action noun, which is just after one entity and the entity is not person and not organization, is 1.

In the following experiments, we investigate the effectiveness of our approaches on under different length limitation of summary. Based on the algorithm of experiment 3, we design experiment to generate summaries with length 50 words, 100 words, 200 words, 400 words. They are named **Experiment 4**, **Experiment 5**, **Experiment 3** and **Experiment 6**.

In other experiments for relevant event-based summarization, we investigate the function of relevance between events. The configurations are described as follows.

Experiment 7: Event terms and event elements are identified as we discussed in Section 3. In this experiment, event elements just include named entities. Occurrences of event terms or event elements are linked with by exact matches. Finally, the PageRank is employed to select important events and then important sentences.

Experiment 8: For reference, we select one of the four model summaries as the final summary for each cluster of documents. ROUGE is employed to evaluate the performance of these manual summaries.

5.3 Experimental Results

The experiment results on independent event-based summarization are shown in table 1. The results for relevant event-based summarization are shown in table 3.

	Exp. 1	Exp. 2	Exp. 3
Rouge-1	0.315	0.322	0.323
Rouge-2	0.049	0.055	0.055
Rouge-L	0.299	0.305	0.306

Table 1. Results on independent event-based summarization (summary with length of 200 words)

From table 1, we can see that results of Experiment 2 are better than those of Experiment 1. It proves our assumption that importance of event terms is different when these event terms occur with different number of event elements. Results of Experiment 3 are not significant better than those of Experiment 2, so it seems that the

assumption that importance of event terms is not very different when these event terms occur with different types of event elements. Another possible explanation is that after adjustment of the weight for event terms, the difference between the results of Experiment 2 and Experiment 3 may be extended.

	Exp. 4	Exp. 5	Exp. 3	Exp. 6
Rouge-1	0.197	0.249	0.323	0.382
Rouge-2	0.021	0.031	0.055	0.081
Rouge-L	0.176	0.231	0.306	0.367

Table 2. Results on independent event-based summarization (summary with different length)

Four experiments of table 2 show that performance of our event based summarization are getting better, when the length of summaries is expanded. One reason is that event based approach prefers sentences with more event terms and more event elements, so the preferred lengths of sentences are longer. While in a short summary, people always condense sentences from original documents, and use some new words to substitute original concepts in documents. Then the Rouge score, which evaluates recall aspect, is not good in our event-based approach. In contrast, if the summaries are longer, people will adopt detail event descriptions in original documents, and so our performance is improved.

	Exp. 7	Exp. 8
Rouge-1	0.325	0.595
Rouge-2	0.060	0.394
Rouge-L	0.305	0.586

Table 3. Results on relevant event-based summarization and a reference experiment (summary with length of 200 words)

In table 3, we found the Rouge-score of relevant event-based summarization (Experiment 7) is better than independent approach (Experiment 1). In Experiment 1, we do not discriminate the weight of event element and event terms. In Experiment 7, we also did not discriminate the weight of event element and event terms. It is fair to compare Experiment 7 with Experiment 1 and it's unfair to compare Experiment 7 with Experiment 3. It looks like the relevance between nodes (event terms or event elements) can help to improve the performance. However, performance of both dependent and independent event-based summarization need to be improved further, compared with human performance in Experiment 8.

6 Discussion

As discussed in Section 2, event-based approaches are also employed in previous works. We evaluate our work in this context. As event-based approaches in this paper are similar with that of Filatovia and Hatzivassiloglou (2004), and the evaluation data set is the same one, the results are compared with theirs.

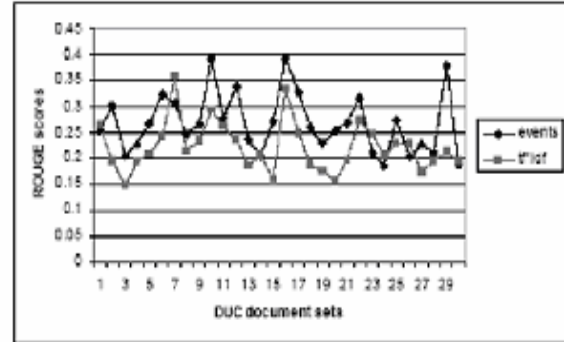


Figure 2. Results reported in (Filatovia and Hatzivassiloglou 2004)

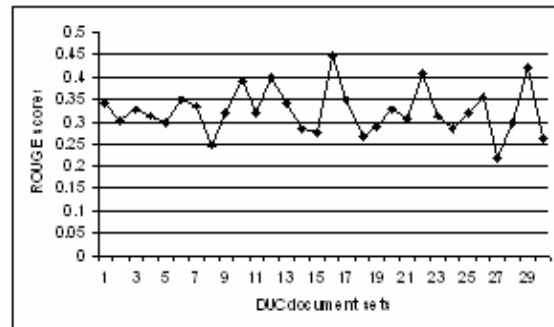


Figure 3. Results of relevant event-based approach

Filatovia and Hatzivassiloglou (2004) report the ROUGE scores according to each cluster of DUC 2001 data collection in Figure 2. In this figure, the bold line represents their event-based approach and the light line refers to tf*idf approach. It can be seen that the event-based approach performs better. The evaluation of the relevant event-based approach presented this paper is shown in Figure 3. The proposed approach achieves significant improvement on most document clusters. The reason seems that the relevance between events is exploited.

Centroid is a successful term-based summarization approach. For comparison, we employ MEAD (Radev et.al., 2004) to generate Centroid-based summaries. Results show that Centroid is better than our relevant event-based approach. After comparing the summaries given by the two approaches, we found some limitation of our approach.

Event-based approach does not work well on documents with rare events. We plan to discriminate the type of documents and apply event-based approach on suitable documents. Our relevant event-based approach is instance-based and too sensitive to number of instances of entities. Concepts seem better to represent meanings of events, as they are really things we care about. In the future, the event map will be build based on concepts and relationships between them. External knowledge may be exploited to refine this concept map.

7 Conclusion

In this study, we investigated generic summarization. An event-based scheme was employed to represent document and identify important content. The independent event-based approach identified important content according to event frequency. We also investigated the different importance of event terms in different context. Experiment showed that this idea achieved promising results. Then we explored summarization under different length limitation. We found that our independent event-based approaches acted well with longer summaries.

In the relevant event-based approach, events were linked together by same or similar event terms and event elements. Experiments showed that the relevance between events can improve the performance of summarization. Compared with close related work, we achieved encouraging improvement.

References

- Regina Barzilay, and Michael Elhadad. 1997. Using lexical chains for text summarization. In Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization, 10-17.
- Ronald Brandow, Karl Mitze, and Lisa F. Rau. 1995. Automatic condensation of electronic publications by sentence selection. *Information Processing and Management* 31(5):675-686.
- John M. Conroy and Judith D. Schlesinger. 2004. Left-brain/right-brain multi-document summarization. Available at <http://duc.nist.gov/pubs.html>
- Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan. 2002. GATE: a framework and graphical development environment for robust NLP tools and applications. In Proceedings of the 40th Annual Meeting of the Association for computational Linguistics (ACL'02).
- Gerald Francis DeJong. 1978. Fast skimming of news stories: the FRUMP system. Ph.D. thesis, Yale University.
- H.P. Edmundson. 1969. New methods in automatic extracting. *Journal of the Association for computing machinery*, 16(2):264-285.
- Elena Filatova and Vasileios Hatzivassiloglou. Event-based extractive summarization. 2004. In Proceedings of the ACL-04 Workshop, 104-111.
- Julian Kupiec, Jan Pedersen and Francine Chen. 1995. A trainable document summarizer. In Proceedings of the 18th ACM-SIGIR conference, 68-73.
- Chin-Yew Lin and Eduard Hovy. 2003. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In Proceedings of HLT-NAACL, Edmonton, Canada, May.
- H.P. Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development* 2:159-165.
- Daniel Marcu. 1997. The rhetorical parsing of natural language texts. In Proceedings of the 35th Annual Meeting of the Association for computational Linguistics (ACL'97), 96-103.
- Dragomir R. Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel, Zhu Zhang. 2004. MEAD - a platform for multidocument multilingual text summarization. LREC 2004.
- Simone Teufel and Marc Moens. 1999. Argumentative classification of extracted sentences as a first step towards flexible abstracting. *Advances in Automatic Text Summarization*, Inderjeet Mani and Mark T. Maybury (editors), 137-154. Cambridge, Massachusetts: MIT Press.
- Larry Page, Sergey Brin, et al. 1998. The PageRank Citation Ranking: Bring Order to the Web. Technical Report, Stanford University, 1998.
- Lucy Vanderwende, Michele Banko, and Arul Menezes. 2004. Event-centric summary generation. Available at <http://duc.nist.gov/pubs.html>
- Masaharu Yoshioka and Makoto Haraguchi. 2004. Multiple news articles summarization based on event reference information. In Working Notes of the Fourth NTCIR Workshop Meeting, National Institute of Informatics, 2004.