ICTGrasper at TAC2009: Temporal Preferred Update Summarization

Jin Zhang
zhangjin@software.ict.ac.cn

Pan Du
dupan@software.ict.ac.cn

Hongbo Xu
hbxu@software.ict.ac.cn

Xueqi Cheng
cxq@ict.ac.cn

Institute of Computing Technology, Chinese Academy of Sciences
NO. 6, South Road, Kexueyuan, Zhongguancun,
Beijing, P.R.China

ABSTRACT

Update summarization is an extension of query-focused multi- document summarization which was launched at DUC 2007. The essential problem of update summarization is to attain the information novelty and topic continuity simultaneously. In this paper, we proposed several Temporal Content Filtering Methods to extract the time-varying information for the update summarization task, while the topic continuity is achieved by identifying the temporal topic signatures. Another manifold ranking approach is also adopted to summarize the topic related information while revealing the intrinsic structure at the same time. The evaluation results show that our approaches are both competitive in practice.

Categories and Subject Descriptors

H.3.1. [**Content Analysis and Indexing**]: Abstracting Methods; I.2.7. [**Natural Language Processing**]: Text Analysis

General Terms

Algorithms, Performance, Experimentation.

Keywords

Update Summarization, Temporal Content Filtering Framework, Manifold Ranking

1. INTRODUCTION

Update Summarization is launched at Document Understanding Conference(DUC)[1] in 2007 as a pilot task. Some changes have been made in Tex Analysis Conference 2008 and 2009. In TAC2009, update summarization aims at generating summaries assuming the user has read some articles before. Specifically, given the topic, the task is to write two summaries, one for document set A and the other for document set B, that address the information need expressed in the corresponding topic statement. The summary for document set A is nothing but a query-focused multi-document summary. The update summary for document set B is also query-focused multi-document one but should be written under the assumption that the user of the summary has already read the documents in document set A. Each summary should be well-organized, in English, using complete sentences. Each summary can be no longer than 100 words.

As an effective and concise approach of helping users to catch the main points, document summarization has attracted much attention since the original work by Luhn[12]. A number of researchers have done good work in multi-document summarization(MDS). Unfortunately, much of their work has focused on the specified static document collection, without attempting to capture the changes over time. Furthermore, the difficulty of constructing an adequate model for dynamically changing information itself is not fully recognized. The classic problem of summarization, simply put, is to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the user's or application's needs[13], which has been studied in many variations and has been addressed through a rich diversity of summarization techniques[5, 7, 16, 20].

The goal of update summarization task is to provide concise, informative summries of the periodical dynamic information devoted to a common topic thus saving the users from browsing the web content during a long time period. We can formulate the update summarization task as dynamic summarization, which can be valuable from periodically monitoring the important changes for the new relevant information over a given time period.

There are several situations when dynamic summarization can be of some value. Users may want to know the most important changes occurring in some domains[9, 1]. They can be interested in popular topics discussed in their area of interest or the changes in public opinions of web pages during a specified period. Additionally, dynamic summarization can also help predict the evolution trend of event in the web. Users can obtain the evolution trend from the sequence of summaries with time going. As a simple application of dynamic summarization, temporal summarization has attracted attention in Topic Detection and Tracking(TDT)[2][17, 18, 21, 3]. As defined in[2], the temporal summarization is in fact a single-document summarization, which is to summarize a single web document over a given time interval. The temporal summarization focuses on the identification of changes between individual web document, however, the challenges of multi-document are seldom addressed.

Research studies on update task of DUC 2007[3] go further by using signature term and term frequency distribution to generate

[1]Document Understanding Conference,http://duc.nist.gov/

[2]TDT,http://www.nist.gov/speech/tests/tdt/

[3]http://www-nlpir.nist.gov/projects/duc/duc2007/tasks.html

TAC 2009, November 16-17, 2009, NIST, Maryland USA
.

summries[4, 8]. In this paper, we introduce a temporal extension of topic signature, called Temporal Topic Signature, to capture the topic continuity of those chronologically ordered document sets. Also, we propose three filtering methods to guarantee the information novelty when modeling the time-varying character of those chronologically ordered document sets. Two ranking approaches-one is signature-based ranking, the other is manifold ranking - are adopted to give each sentence of each document set a ranking value measuring the salience of the each sentence. Then those most salient sentences are extracted as components of summaries. Experiments are conducted to evaluate the effectiveness of our system, and the results on TAC 2008 data set show that our approaches are competitive with state-of-the-art systems developed in this area.

In Section 2, we give an overview of our summarization system. The Content filtering methods and demonstrated in section 3. The ranking approaches we adopted are introduced in section 4, with experiments and evaluation followed in section 5. Finally, we conclude this paper with a summary and discussion of results in TAC 2009, and look ahead to future work.

2. SYSTEM OVERVIEW

In this section,we present the overview of our summarization system on TAC2009 Update Summarization Track. The system architecture is shown as figure 1:

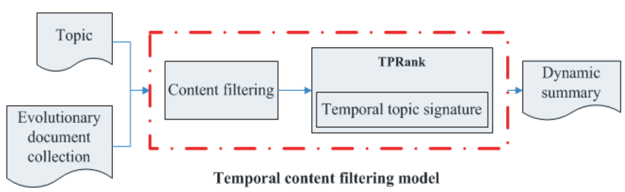

Figure 1: TeCFilM Framework

The Content Filtering module is in charge of identifying the time-varying information from the document collections. We proposed several filtering models including Document Filtering Model, Union Filtering Model and Summary Filtering Model to achieve this goal.

The duty of Summary Generation module is to create the final summary by selecting proper sentences from the primitive document set. Based on the Temporal Signature Term we proposed, not only the salient information of certain topic, but also the novel information can be identified simultaneously.

We also adopted a manifold ranking approach to generate the corresponding summary.

3. CONTENT FILTERING METHODS

To capture the changed information of the document collections, the first challenge is to filter the redundant information from current documents comparing to the history documents. We adopt the degree of membership from fuzzy set to measure the similarity of sentences between history information and current information. In our practical method, we simply take the sentences of the last collection D_n as the current information I_c, while all those previous collections as the history information I_h. Correspondingly, the key issue of the update summarization can be simplified as determining the time-varying information from I_c to I_h. Contents Filtering Methods introduces the fuzzy conjunctive operator to identify the redundancy of sentences between sets A and B by borrowing the degree of membership from fuzzy set theory:

$$A\tilde{\cap}B = \{s|similarity(s, s_k) \geq m_d, s \in A, s_k \in B\} \quad (1)$$

where $m_d \in [0, 1]$ is degree of membership. The filtering operation between sets A and B can be defined as follows:

$$A - B = \{s|s \in A, \{s\} \nsubseteq A\tilde{\cap}B\} \quad (2)$$

Given the history information I_h, the current information I_c, the degree of membership m_d, and the summarization function f, there are totally three filtering methods for update summarization according to the objects to be filtered.

3.1 Document Filtering Method

The first content filtering method is document filtering method (DFM), where the object to be filtered is document collection of current information I_c itself. In the method, we assume that the time-varying content is the sentences of I_c except the sentences redundant to I_h, and these redundant sentences can be filtered from I_c with a specified degree.

The document filtering process can be conducted by computing I_c - I_h following the filtering operation defined above.

In general, the summary of a document collection has high relevant to its content. Thus, in order to save the calculating cost of redundant content, the map of I_h in summary space, $f(I_h)$, is used to substitute I_h, then we can obtain an alternative method, $I_c - f(I_h)$. In order to discriminate these two varieties, $I_c - I_h$ and $I_c - f(I_h)$ are denoted by *DFM1* and *DFM2* respectively.

3.2 Summary Filtering Method

The second content filtering method is summary filtering method (SFM), where the object to be filtered is the summary of current information, denoted by $f(I_c)$.

In SFM, we assume that the update summary can be generated by filtering the redundant sentences from $f(I_c)$ according to I_h. Intuitively, there are fewer sentences in summary, thus the calculating cost of redundant content of SFM is lower than that of DFM.

Similarly, in order to further save filtering cost, I_h can be substituted with $f(I_h)$, then two varieties of SFM can be obtained. For convenience, *SFM1* and *SFM2* are used to denote $f(I_c) - I_h$ and $f(I_c) - f(I_h)$.

3.3 Union Filtering Method

With the assumption that the relation between history information and current information cannot be omitted, the update summary can be generated from the union of I_h and I_c, in that the third content filtering method can be presented as union filtering method (UFM), where the object of content filtering is the summary of the union of I_h and I_c.

In like manner, two varieties of UFM, $f(I_c + I_h) - I_h$ and $f(I_c + I_h) - f(I_h)$ can be obtained. Here, we use *UFM1* and *UFM2* to denote them respectively.

4. RANKING APPROACHES

We directly select sentences from the document set to generate the summary according to their ranks we got. So ranking approaches is the kernel components of our summarization system. Totally, we adopted two ranking strategies, one is the Temporal Preferred Ranking, which can capture not only the salience of sentences but also their novelty, the other is the manifold ranking, which is good at capturing the intrinsic structure of the document sets if we take each sentence as a point in the non-Euclidean document space.

4.1 Temporal Preferred Ranking

Although the filtered sentences may be relevant to the specified topic, they do not necessarily convey the equivalent information with each other.

As first proposed in [11], the topic of document collection can be represented using a set of terms - known as topic signatures (TS) - that are highly correlated to the topic itself. They took the co-occurrence relationship into consideration to expand topic representation. Inspired by this method, we employ sets of extracted phrases ordered chronologically, ie., temporal topic signatures (TTS), to represent the relevance of the sentences to the topic with time going. The definition of TTS is as follows:

$$\begin{aligned} TTS &= \{topic, signature^T\} \\ &= \{topic, \langle sig^1, \ldots, sig^i, \ldots, sig^n\rangle\} \end{aligned} \tag{3}$$

where,

$$\begin{aligned} &sig^i = \langle (t_1^i, score_1^i), \ldots, (t_j^i, score_j^i), \ldots, (t_s^i, score_s^i)\rangle, \\ &i \in \{1, \ldots, n\}. \end{aligned} \tag{4}$$

From the definition we can see, a TTS has multiple sub-signatures, denoted as sig^i, ordered chronologically with respect to one single topic, which is different from TS. Sub-signature sig^i of interval i is extracted from the temporal document collection D_i of the same time interval. Each sub-signature is a term-score vector extracted from temporal document collection, for example, t_j^i with $score_j^i$ is the j_{th} element of vector sig^i extracted from D_i.

Usually, the weight of a term is measured by term frequency (*TF*), yet the frequent-usage of a term does not guarantee that it is a meaningful term, and a meaningless but important term will undermine the summary precision. A meaningful term can be guaranteed by its independent usage, eg. context independency (*AV*) [6]. We employ both Accessor Variety (*AV*) and Term Frequency (*TF*) to extract each sub-signature.

In order to evaluate the significance of the terms quantitatively from different perspectives, we employ a linear combination of AV and TF as the hybrid score of term t.

$$score(t) = AV(t) + \lambda * TF(t) \tag{5}$$

The sub-signature is then acquired by selecting the top terms ranked by calculated scores. Being composed of independent sub-signatures, *TTS* can be obtained after all its sub-signatures are extracted.

A temporal preferred ranking score of a sentence can be represented with two aspects, the topic relevant score and the temporal score of sentence measured by temporal topic signatures.In this paper, we adapt a linear combination to measure the temporal preferred scores of sentences as follows:

$$Score_{tp}(s) = \alpha * SigRel(s) + (1 - \alpha) * SigTemp(s) \tag{6}$$

where s is a sentence, $Score_{tp}(s)$ is the temporal preferred score of s, $SigRel(s)$ and $SigTemp(s)$ are the topic preferred score and the temporal score of s respectively. In the linear combination, α is a weight factor. Given the temporal topic signature, the temporal score $SigTemp(s)$ of a specified sentence s is computed as:

$$SigTemp(s) = \sum_{i=1}^{k} score(t_i) \tag{7}$$

where t_1,...,t_k are the terms occur both in the sentence s and the sub-signature of the document set which s belongs to, and $score(t_i)$ is the normalized score of t_i determined by AV and TF.

TPRank can be formulated as the following,

$$\overrightarrow{f(tp)} = \alpha * p'_{tp} + (1 - \alpha) * M * \overrightarrow{f(tp)}' \tag{8}$$

where $\overrightarrow{f(tp)}$ is the vector of temporal preferred ranking values for sentences, and p_{tp} is the normalized vector of temporal preferred scores generated by Equation 7. Empirically, the "damping factor" α between 0.1 and 0.2 is more significant for rank criteria [5].

4.2 Manifold Ranking

Another ranking approach is Manifold Ranking[23, 22]. This is a ranking approach based on semi-supervised learning, the goal of it is to rank the data with respect to the intrinsic global manifold structure collectively revealed by a huge amount of data. Generally speaking, for many real world data types this would be superior to a local method, which rank data simply by pairwise Euclidean distances or inner products.

This approach was successfully adopted by Wan et al.[19] to deal with topic-focused multi-document summarization. The prior assumption of manifold ranking is: (1) nearby points are likely to have the similar ranking scores; (2) points on the same structure (typically referred to as a cluster or a manifold) are likely to have the same ranking scores.

In the summarization context, the data points are denoted by topic description and all sentences in the documents. The ranking process can be formalized as follows:

Given a set of points $\chi = \{x_0, x_1, ..., x_n\} \subset \mathbb{R}^m$, the first points is the topic description or a query and the rest are the points that we want to rank according to their relevance to the topic or query. Let $f : \chi \longrightarrow \mathbb{R}$ denote a ranking function which assigns to each point x_i a ranking value f_i. We can view f as a vector $f = [f_0, ..., f_n]^T$. We also define the vector $y = [y_0, ..., y_n]^T$, in which $y_0 = 1$ because x_0 is a query or topic description, and $y_i = 0 (1 \leq i \leq n)$ for all the sentences in the document set. The ranking algorithm is shown as figure 2. The normalization in the third step guarantee convergence of the algorithm.

1. Compute the pair-wise similarity values between sentences using the standard Cosine measure. Given two sentences x_i and x_j, the Cosine similarity is denoted as $sim(x_i, x_j)$, computed as the normalized inner product of the corresponding term vectors.

2. Connect any two points with an edge if their value exceeds 0. We define the affinity matrix W by $W_{ij} = sim(x_i, x_j)$ if there is an edge linking x_i and x_j. Note that we let $W_{ii} = 0$ to avoid loops in the graph.

3. Symmetrically normalized W by $S = D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with (i, i)-element equal to the sum of the i-th row of W.

4. Iterate $f(t+1) = \beta S f(t) + (1-\beta)y$ until convergence, where $0 < \beta < 1$.

5. Let f_i^* denote the limit of the sequence $f_i(t)$. Each sentences $x_i (1 \leq i \leq n)$ gets its ranking score f_i^*.

Figure 2: The manifold ranking algorithm.

The iteration in the fourth step is the key step for all points to spread their ranking scores to their neighbors via the weighted network. The parameter of manifold-ranking weight β specifies the relative contributions to the ranking scores from neighbors and the initial ranking scores. Note that self-reinforcement is avoided since diagonal elements of the affinity matrix set to zero. The sequence $f(t)$ converges to

$$f^* = \gamma(I - \beta S)^{-1} y \tag{9}$$

where $\gamma = 1 - \beta$. This is proven by Zhou et al.[22].

5. EXPERIMENTS

5.1 Data Set

Update summarization has been evaluated on DUC 2007 and TAC 2008, each task having a gold standard data set consisting of document clusters and reference summaries. The test data set of TAC 2009 is composed of 44 topics. Each topic has a topic statement (title and narrative) and 20 relevant documents which have been divided into 2 sets: Document Set A and Document Set B. Each document set has 10 documents, and all the documents in Set A chronologically precede the documents in Set B.

In our experiments, data set from TAC 2008 including the two chronologically ordered document sets A and B, together with their corresponding human model summaries are used for training and parameter tuning, data set from TAC 2009 without gold standard models are used for testing, and the results are evaluated and published by organizer of TAC 2009. As a preprocessing step, the stop words in each sentence were removed and the remaining words are stemmed.

5.2 Evaluation Metric

ROUGE[10], Recall Oriented Understudy for Gisting Evaluation, is a metric adopted by TAC for automatically summarization evaluation. There are several variants can be used in practice with provided toolkits. ROUGE-N measures summary quality by counting overlapping units of n-gram between the candidate summary(peer) and the reference summaries(model). The evaluation metrics we adopted in our training process are ROUGE-1, ROUGE-2 and ROUGE-SU4 respectively. ROUGE-N is computed as follows:

$$ROUGE - N = \frac{\sum_{S \in ReferenceSummries} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in ReferenceSummries} \sum_{gram_n \in S} Count(gram_n)} \quad (10)$$

Where n stands for the length of n-gram, $gram_n$, and $Count_{match}(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries. ROUGE-SU4 is a skip-bigram co-occurrence measure with addition of unigrams as counting unit.

The ROUGE toolkit reports scores for 1-, 2-, 3-, and 4-gram. We show three of the ROUGE metrics in the experimental results, at a confidence level of 95%: ROUGE-1, ROUGE-2, ROUGE-SU4.

Pyramid[14, 15] is a manual metric used for summary evaluation in TAC 2009. Its kernel concept is Summary Content Units, referred as SCUs, which are semantically motivated, sub-sentential units that are variable in length but no bigger than a sentential clause. SCUs emerge from annotation of a collection of human summaries for the same input. They are identified by noting information that is repeated across summaries, whether the repetition is as small as a modifier of a noun phrase or as large as a clause. The weight an SCU obtains is directly proportional to the number of reference summaries that support that piece of information. The evaluation method that is based on overlapping SCUs in human and automatic summaries is described in the Pyramid method.

5.3 Experimental Results

5.3.1 Parameter Tuning

Figure 3 demonstrates the influence of parameter α in our temporal preferred ranking algorithm. All the ROUGE scores, including ROUGE-1, ROUGE-2, and ROUGE-SU4, arrive at their maximum value approximately when $\alpha = 0.3$. The parameter was trained on the data set of TAC 2008, update summarization task.

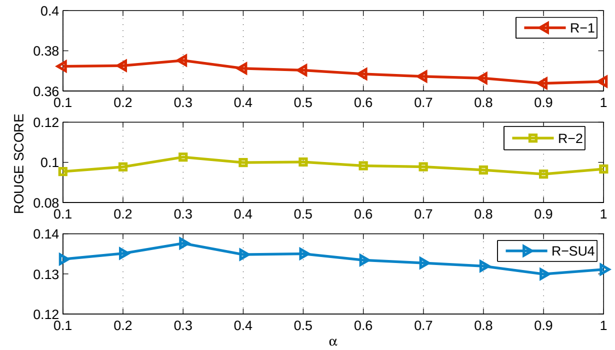

Figure 3: ROUGE Scores vs. α on Data Set of TAC2008.

Since α is used to balancing the influence between topic relevance and temporal topic nature, it exposes that the temporal topic nature show more importance than topic relevance, making the final score strongly reflect the update nature of the generated summaries. The three scores of ROUGE-1, ROUGE-2, and ROUGE-SU4 agree well on the parameter α.

Figure 4 demonstrates the influence of parameter β in out manifold ranking algorithm. Three ROUGE scores - ROUGE-1, ROUGE-2, ROUGE-SU4 - are obtained in our training process conducted on data set of TAC 2008, update summarization task.

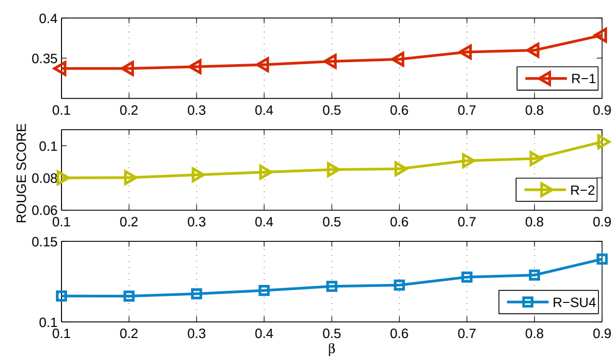

Figure 4: ROUGE Scores vs. β on Data Set of TAC2008.

All the three scores arrive at their climax near the point of $\beta = 0.9$. The high value of β means that the scores transposed to the point's neighbors from the prior score of its own at each iteration step are very low $(1 - \beta)y$, while much of the score transposed to their neighbors is the score that it has cumulated from the iteration process.

5.3.2 System Comparison

Our proposed approaches for update summarization has accomplished a competitive performance on Text Analysis Conference of 2009. The update summarization task on TAC 2009 requires to generate 100-word summaries for 44 topics. Each topic has a topic statement and 20 relevant documents which have been divided into 2 sets: Document Set A and Document Set B. Each document set has 10 documents, and all the documents in Set A chronologically precede the documents in Set B. The generated summaries are evaluated by National Institute of Standards and Technology (NIST[4]). All summaries were first truncated to 100 words to be identified for automatic evaluation of ROUGE Metrics. The evaluation results of temporal preferred system of Run 4 are demonstrated as Table 1:

Table 1: The Evaluation Results of Run 4 on TAC 2009.

Metric	Score	Rank
Pyramid - A	0.320	6
Pyramid - B	0.290	6
BE - A	0.05834	4
BE - B	0.05478	6
ROUGE-2 - A	0.10353	7
ROUGE-2 - B	0.09138	4
ROUGE-SU4 - A	0.13878	7
ROUGE-SU4 - B	0.13331	6

Except for the automatic metrics of ROUGE family, NIST also conducted a manual evaluation of summary content based on the Pyramid Method[5], in which each topic statement and its 2 document sets were given to 4 different NIST assessors.

The evaluation results of manifold ranking based approach - Run 45 on TAC 2009 - are demonstrated in Table 2:

Table 2: The Evaluation Results of Run 45 on TAC 2009.

Metric	Score	Rank
Pyramid - A	0.332	4
Pyramid - B	0.292	5
BE - A	0.05894	3
BE - B	0.05021	8
ROUGE-2 - A	0.10637	4
ROUGE-2 - B	0.08520	10
ROUGE-SU4 - A	0.13990	5
ROUGE-SU4 - B	0.12582	9

Among the total 52 runs from 27 participants for the update summarization task, although competitive of systems are, there still much work to be done to enhance them. Comparing to other participants, the performance of our system 4 is rather stable on ranks of both sets of A and B, which means that it captured the update nature of set B successfully. However, system of run 45 only achieved a promising performance on set A. The poor performance on set B of system 45 shows its weakness of capturing the update nature of information when doing summarization. The results on set A of system 45, on the other hand are even better performed than system 4, which means that system 45 are more good at query-focused multi-document summarization than system 4. As can be seen in Figure 5.

Upon these analysis, we believe that by combining the well-performed side of the systems of 4 and 45, a promising performance can be expected in our future work.

6. CONCLUSIONS AND FUTURE WORK

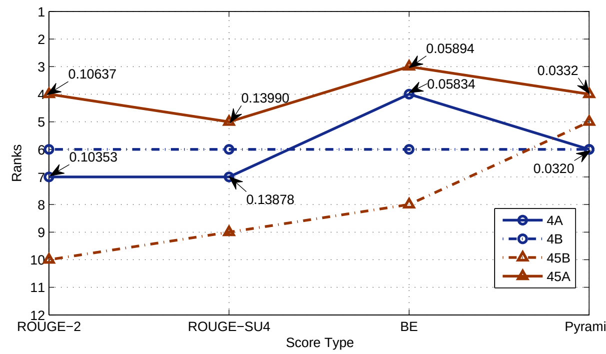

Figure 5: Result Analysis of our systems on TAC 2009.

In this paper, we proposed two approaches for update summarization task of TAC 2009. The first one is a signature based approach, in which temporal topic signatures are extracted after the information of the chronologically organized documents have been filtered through three filtering strategies. In this way, the information can hold their topic continuity and information novelty simultaneously. The second one is manifold ranking based approach, in which the macro-structure of the information can be reserved, reflecting a better relevance of the query or topic.

Both of our approaches have achieved promising results on TAC 2009 under of the evaluation metrics of ROUGE, BE and Pyramid. The first approach is more good at capture the update essence of the information according to the evaluation results provided by NIST, what's more, it achieved a stable performance on every evaluation metric. While the second approach does the query-focused multi-document summarization better than the first one, as can been from Figure 5, although it fails to capture the update nature of information properly.

We will consider the combination of these two approaches together, by adding the power of query-focused multi-document summarization of system 45 and the ability of capturing updated information of system 4, we believe a better performance can be achieved in our future work.

7. ACKNOWLEDGMENTS

This work is supported by National Science Foundation of China under Grant No. 60903139 and National Hi-Tech Research and Development Program of China under Grant No. 2007AA01Z441.

8. ADDITIONAL AUTHORS

9. REFERENCES

[1] J. A. and I. M. Temporal multi-page summarization. *Web Intelli. and Agent Sys.*, 4(2):163–180, 2006.

[2] J. Allan, R. Gupta, and V. Khandelwal. Temporal summaries of news topics. In *Proceedings of SIGIR'01*, pages 10–18, New Orleans, Louisiana, United States, 2001.

[3] B. Cao, D. Shen, J.-T. Sun, X. Wang, Q. Yang, and Z. Chen. Latent factor detection and tracking with online nonnegative matrix factorization. In *Proceedings of IJCAI'07*, pages 2689–2694, Hyderabad, India, 2007.

[4] http://www.nist.gov/

[5] http://www1.cs.columbia.edu/~becky/DUC2006/2006-pyramid-guidelines.html

[4] J. Conroy, J. Schlesinger, and D. O'Leary. Classy 2007 at duc 2007. In *Proceedings of Document Understanding Conference*, 2007.

[5] G. Erkan and D. R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res. (JAIR)*, 22:457–479, 2004.

[6] H. Feng, K. Chen, X. Deng, and W. Zheng. Accessor variety criteria for chinese word extraction. *Computational Linguistics*, 30:75–93, 2004.

[7] S. M. Harabagiu and V. F. Lacatusu. Topic themes for multi-document summarization. In *Proceedings of SIGIR'05*, pages 202–209, Salvador, Brazil, 2005.

[8] A. Hickl, K. Roberts, and F. Lacatusu. Lcc's gistexter at duc 2007: Machine reading for update summarization. In *Proceedings of DUC'07*, 2007.

[9] A. Jatowt, K. Khoo, and M. Ishizuka. Change summarization in web collections. In *Proceedings of IEA/AIE'04*, pages 653–662, 2004.

[10] C.-Y. Lin. Rouge: a package for automatic evaluation of summaries. In *Proceedings of Workshop on Text Summarization*, Barcelona, Spain, 2004.

[11] C.-Y. Lin and E. Hovy. The automated acquisition of topic signatures for text summarization. In *Proceedings of the COLING'00*, pages 495–501, Saarbrücken, Germany, 2000.

[12] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal*, pages 159–165, 1958.

[13] I. Mani. Recent developments in text summarization. In *Proceedings of CIKM'01*, pages 529–531, Atlanta, Georgia, USA, 2001.

[14] A. Nenkova, R. Passonneau, and K. McKeown. The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Trans. Speech Lang. Process.*, 4(2):4, 2007.

[15] A. Nenkova and R. J. Passonneau. Evaluating content selection in summarization: The pyramid method. In *HLT-NAACL*, pages 145–152, 2004.

[16] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen. Document summarization using conditional random fields. In *Proceedings of IJCAI'07*, pages 2862–2867, Hyderabad, India, 2007.

[17] R. Swan and J. Allan. On-line new event detection and tracking. In *Proceedings of SIGIR'98*, pages 37–45, Melbourne, Australia, 1998.

[18] R. Swan and J. Allan. Automatic generation of overview timelines. In *Proceedings of SIGIR'00*, pages 49–56, Athens, Greece, 2000.

[19] X. Wan, J. Yang, and J. Xiao. Manifold-ranking based topic-focused multi-document summarization. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2903–2908, Hyderabad, India, January 6-12 2007.

[20] J. Zhang, X. Cheng, and H. Xu. Dynamic summarization: Another stride towards summarization. In *Proceedings of WI-IAT'07*, pages 64–67, Silicon Valley, CA, 2007.

[21] Y. Zhang, J. Callan, and T. Minka. Novelty and redundancy detection in adaptive filtering. In *Proceedings of SIGIR'02*, pages 81–88, 2002.

[22] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[23] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.