NTT's Multiple Document Summarization System for DUC2004

Tsutomu HIRAO, Jun SUZUKI, Hideki ISOZAKI and **Eisaku MAEDA**
NTT Communication Science Laboratories, NTT Corp.
2-4,Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan
{hirao,jun,isozaki,maeda}@cslab.kecl.ntt.co.jp

Table 1: cross tabulation

	T	$\neg T$
p	n_{11}	n_{12}
$\neg p$	n_{21}	n_{22}

Abstract

We participated in the Document Understanding Conference 2004 (DUC 2004) to confirm the effectiveness of our multiple document summarization system which uses a sequential-pattern mining technique.

1 Introduction

In this paper, we provide a description of our system for summarizing multiple documents. Our system employs a sequential pattern mining algorithm (PrefixSpan)(Pei et al., 2001) for sentence extraction and uses Maximum Marginal Relevance (Carbonell and Goldstein, 1998; Goldstein et al., 2000) to minimize the redundancy of extracted sentences.

The results of task-2 at DUC2004 revealed that our system need improvement enough.

2 Sentence Extraction Phase

Conventional summarization methods utilize a TF·IDF model for significance scores of sentences (Zechner, 1996). Lin (Lin and Hovy, 2000) extended such methods by proposing a method based on not only unigrams but also n-grams. However, these methods ignore gappy n-grams. Therefore, we use not only n-grams but also gappy n-grams by using a sequential-pattern mining method.

2.1 Sequential Pattern Extraction

We can extract a sequential pattern, *i.e.,* both n-grams and gappy n-grams from text by using a text mining algorithm, PrefixSpan (Pei et al., 2001). However, extracted patterns are not always effective. Therefore, we identify the significant patterns for sentence extraction from a given document set by using χ^2 test.

For each pattern p in document set T, we make a cross-tabulation list (Table 1). n_{11} indicates the number of sentences that contain t in document set T, and n_{12} indicates the number of sentences that contain t expect for document set T. n_{21} indicates the number of sentences that do not contain t in document set T, and n_{22} indicates the number of sentences that do not contain t expect for document set T.

Here, χ^2 metrics is defined as follows:

$$\chi^2 = \frac{(n_{11}+n_{12}+n_{21}+n_{22})(n_{11}n_{22}-n_{12}n_{21})^2}{(n_{11}+n_{12})(n_{11}+n_{21})(n_{12}+n_{22})(n_{21}+n_{22})} \quad (1)$$

We used the top 1,000 patterns for sentence scoring. Table 2 shows examples of the sequential pattern extracted by PrefixSpan with χ^2 metric.

2.2 Sentence Scoring

We define the weight of a pattern as follows:

$$w(p) = \frac{\log(f(p,T)+1)\cdot\log(\frac{|DB|}{f(p,DB)})}{len(p)}. \quad (2)$$

$f(p,T)$ is the sentence frequency of pattern p in the document set, and $f(p,DB)$ is the sentence frequency of pattern p in all topics. $|DB|$ is the number of sentences in all topics, and $len(p)$ is the length of the pattern.

Table 2: Examples of sequential pattern (topic-id = d30026t).

Netscape	America Online	America Online be
AOL	Netscape software	be America Online
Online	Netscape be	America Online Internet
software	be commerce	America Online Netscape
Microsoft	AOL Netscape	America Online have
Internet	Netscape Sun	America Online service
Sun	be software	America Online online
commerce	Netscape Internet	America Online company
America	AOL Sun	America Online say
service	Online be	have America Online
online	be Netscape	America Online software
company	be Internet	be online store
Web	be service	AOL Netscape software
store	be Online	AOL Netscape Sun
computer	Online Internet	AOL Sun Microsystems
sell	America Internet	Netscape software commerce
technology	Sun be	say America Online
deal	commerce service	Netscape business software
market	AOL software	Netscape business Internet
Microsystems	be company	Netscape have commerce

$A = \{\}$;
$R = \{S_1, S_2, \cdots, S_\ell\}$;
N =The number of sentences as output;
While($|A| < N$){

$S^* = \mathrm{MMR}(A, R)$;
$A = A \cup \{S^*\}$;
$R = R - \{S^*\}$;

}
Output A, where

$$\mathrm{MMR}(A,R) = \begin{cases} \underset{S_i \in R}{\operatorname{argmax}}\, s(g(S_i)) & \text{if} \quad A = \phi \\ \underset{S_i \in R}{\operatorname{argmax}}\, (\alpha score(S_i) & - \\ \quad (1-\alpha) \max_{S_j \in A} \mathrm{Sim}(S_i, S_j)\big) \end{cases}$$

Figure 1: Reordering algorithm by MMR

Finally, we define the sentence score as follows:

$$\mathrm{score}(S_i) = \sum_{p \in S_i} w(p). \tag{3}$$

3 Redundancy Minimization Phase

It is said that a document set includes redundant sentences. To minimize redundancy, Carbonell proposed Maximum Marginal Relevance (MMR) (Carbonell and Goldstein, 1998; Goldstein et al., 2000). MMR deals with two factors: a significance score of a sentence and the similarity between the sentence and sentences already selected for summary.

Figure 1 shows a reordering algorithm based on MMR. In the figure, R is the set of all sentences in a given document set. A is the set of sentences selected for summary. $\mathrm{Sim}(S_i, S_j)$ provides the similarity between sentence S_i and sentence S_j, and α is a trade-off parameter for the two arguments. We set α as 0.6 in our submission.

Here, we use Word Sequence Kernel (WSK) (Cancedda et al., 2003) as the similarity between sentences because WSK can measure the similarity considering sequential patterns.

4 Results

We describe the evaluation results of task-2 in DUC2004. Table 3 shows the results of using an automatic evaluation method called ROUGE. Our system's ID is "123." Our system ranked almost 32nd out of all systems. On the other hand, Table 4 shows both content and readability evaluation results by human subjects. In the table, Cov is "mean coverage," and a high score represents a good performance; Q1 - Q7 are "count of quality questions," and a low score means a good performance.

In the case of automatic evaluation, our system was outperformed by systems 27, 138 and 117. However, subjective evaluation showed that our system

Table 3: Evaluation results by ROUGE

Method	rouge-1	rouge-2	rouge-3	rouge-4	rouge-L	rouge-W
Ave. score	0.285	0.0485	0.0126	0.00488	0.305	0.104

Table 4: Evaluation results by human subjects

ID	Cov	Q1	Q2	Q3	Q4	Q5	Q6	Q7
HUMAN	0.445	1.780	1.580	1.060	1.500	1.020	1.380	1.400
65	0.303	2.860	2.520	1.620	1.600	1.440	1.400	1.340
124	0.262	2.820	2.560	2.260	1.640	1.580	1.420	1.440
44	0.262	2.560	2.360	1.580	1.600	1.180	1.300	1.320
93	0.255	2.980	2.520	1.700	1.460	1.380	1.540	2.700
81	0.247	2.760	2.740	1.660	1.920	1.300	1.380	1.340
55	0.243	3.060	2.680	1.400	2.380	1.180	1.580	1.420
120	0.243	2.320	2.080	1.560	1.200	1.460	1.220	1.380
102	0.242	2.680	2.660	1.520	1.640	1.200	1.420	1.400
19	0.224	3.220	2.840	1.460	2.540	1.020	1.620	1.700
34	0.222	3.240	2.580	1.520	2.420	1.140	1.380	1.520
11	0.216	2.600	2.600	2.180	1.420	1.620	1.220	1.200
2	0.200	1.440	2.260	1.340	1.300	1.020	1.300	1.320
123	0.170	3.340	2.840	1.660	2.760	1.160	1.280	1.460
27	0.166	3.280	2.700	1.360	2.340	1.080	1.220	1.360
138	0.165	2.420	2.680	1.760	1.540	1.080	3.360	2.520
117	0.115	4.820	4.540	2.060	4.240	1.500	4.300	2.000
111	0.049	4.740	4.440	1.400	4.660	1.120	1.680	2.300

outperformed them. These results indicate that automatic evaluation includes some error.

5 Conclusion

We described our system, which is based on sequential pattern mining and MMR, and our participation in the multiple document summarization tasks at DUC 2004. We also provided evaluation results.

Acknowledgements

We would like to thank Dr. Sekine for allowing us to use his English sentence analyzer OAK System.

References

N. Cancedda, E. Gaussier, C. Goutte, and J-M. Renders. 2003. Word-sequence kernels. *Journal of Machine Learning Research*, 3(Feb):1059–1082.

J. Carbonell and J Goldstein. 1998. The use of mmr, diversity-based reranking for reordering document and producing summaries. *Proc. of the 21th Annual International ACM SIGIR Conference on Research and Development in the Information Retrieval*, pages 335–336.

J. Goldstein, V. Mittal, J. Carbonell, and J. Callan. 2000. Creating and evaluating multi-document sentence extract summaries. *Proc. of the 9th International Conference on Information and Knowledge Management*, pages 165–172.

C-Y Lin and E.H. Hovy. 2000. The automated acquisition of topic signatures for text summarization. *Proc. of the 18th International Conference on Computational Linguistics*, pages 495–501.

J. Pei, J. Han, B. Mortazavi-Asl, and H. Pinto. 2001. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. *Proc. of 17th International Conference on Data Engineering (ICDE 2001)*, pages 215–224.

K. Zechner. 1996. Fast generation of abstracts from general domain text corpora by extracting relevant sentences. *Proc. of the 16th International Conference on Computational Linguistics*, pages 986–989.