# TREC-7 Experiments: Query Expansion Method Based on Word Contribution

Keiichiro Hoashi, Kazunori Matsumoto, Naomi Inoue, and Kazuo Hashimoto
{hoashi,matsu,inoue,kh}@kddlabs.co.jp

KDD R&D Laboratories, Inc.
2-1-15 Ohara Kamifukuoka
Saitama 356-8502 JAPAN

## 1 Introduction

This is KDD R&D Laboratories' first participation in TREC. In this participation, we focused on experiments on a novel method of query expansion.

The query expansion method described in this paper is based on a measure we call "word contribution". Word contribution is a measure which expresses the influence of a word to the similarity between the query and a document. From our data, we figured that words which have highly negative contribution can be considered as to being expressive of the characteristics of the data (query or document) in which they exist. We proposed extracting such words from documents highly similar to a query, and adding them to the original query to generate an expanded query. We made experiments to evaluate this method, and reported the results in this paper.

We submitted 3 official ad hoc runs (**KD70000**, **KD71010q**, **KD71010s**) to TREC-7. However, the data we used for these runs were generated by a buggy morphological analysis program, which we consider a serious cause for our bad results. Since the official submission, we have fixed these bugs, and reconstructured our data. The results described in this paper are based on these new data, and some experiments made after the TREC-7 conference.

## 2 Retrieval Method

### 2.1 Indexing

For indexing the topics and the documents, we ran a morphological analysis program on the data, and extracted nouns, proper nouns, and undefined words. A frequency table was made for each datum consisted of the extracted terms and frequency. The morphologial analysis program was the program in which bugs were discovered after the official submission.

In our experiments, we used the data from the TREC CD-ROMs 4 and 5, excluding the Congressional Reports. The total number of terms extracted from this data was 772,659.

### 2.2 Similarity Calculation

For similarity calculation, we applied a probabilistic model proposed by Iwayama et al [1]. This model is based on an idea called *Single random Variable with Multiple Values* (SVMV), and was proved effective in text categorization compared to other existing methods.

The formula for similarity calculation between documents $d_1$ and $d_2$ for SVMV is described in Figure 1, where:

| | |
|---|---|
| $Sim(d_1, d_2)$: | Similarity of documents $d_1$ and $d_2$ |
| $F_d(d, w)$: | Frequency of word $w$ in document $d$ |
| $N_d(d)$: | Number of words in document $d$ |
| $F(w)$: | Frequency of word $w$ in all documents |
| $N$: | Number of words in all documents |

## 3 Query Expansion Based on Word Contribution

In this section, we will make an explanation of our proposed method of query expansion based on word contribution.

$$Sim(d_1, d_2) = M(d_1, d_2) - U(d_1) - U(d_2)$$

$$U(d) = \log \left( \sum_{w \in d} \frac{\left(\frac{F_d(d,w)}{N_d(d)}\right)^2}{\frac{F(w)}{N}} \right)$$

$$M(d_i, d_j) =$$

$$\sum_{d \in d_i, d_j} \log \left( \sum_{w \in d} \frac{\frac{F_d(d,w)}{N_d(d)} \cdot \frac{F_d(d_i,w)+F_d(d_j,w)}{N_d(d_i)+N_d(d_j)}}{\frac{F(w)}{N}} \right)$$

Figure 1: Similarity calculation formula for SVMV.

### 3.1 Definition of Word Contribution

Word contribution is a measure which expresses the influence of a word (or term) to the similarity between the query and a document. It is defined by the following formula:

$$Cont(w, q, d) = Sim(q, d) - Sim(q'(w), d'(w))$$

where $Cont(w, q, d)$ is the contribution of the word $w$ in the similarity between query $q$ and document $d$, $Sim(q, d)$ is the similarity between $q$ and $d$, $q'(w)$ is query $q$ excluding word $w$, and $d'(w)$ is document $d$ excluding word $w$. In other words, the contribution of word $w$ is the difference between the similarity of $q$ and $d$, and the similarity of $q$ and $d$ when word $w$ is assumed to be non-existent in both data. Therefore, there are words which have positive contribution, and words which have negative contribution. Words with positive contribution increases the similarity, and words with negative contribution decreases the similarity. An example of word contribution data calculated from TREC-6 data is shown in Table 1.

### 3.2 Hypothesis

Figure 2 illustrates the contribution of all words from the query and document used in the example shown in Table 1. The document used in this example is relevant to the query. The data is sorted in descending order according to the contribution of each word.

From Figure 2, it is apparent that there are only a small number of words with highly positive contribution, and a small number of words

Table 1: Words with 10 highest/lowest contribution in Topic 313 and FT932-6259

| Word | Contribution |
|---|---|
| levitation | 0.39429400 |
| discussion | 0.00030683 |
| plan | 0.00012887 |
| year | -1.33E-06 |
| government | -5.21E-06 |
| system | -0.0000053 |
| city | -5.77E-06 |
| take | -6.54E-06 |
| development | -6.92E-06 |
| agreement | -7.91E-06 |
| ... | ... |
| narrative | -0.0009595 |
| JAPANESE | -0.0009871 |
| JFK | -0.0044444 |
| Guardia | -0.0046405 |
| superconductivity | -0.0107779 |
| Nakamoto | -0.0114731 |
| Michiyo | -0.0137173 |
| flywheel | -0.039495 |
| Grumman | -0.0424242 |
| motor_car | -0.1363256 |

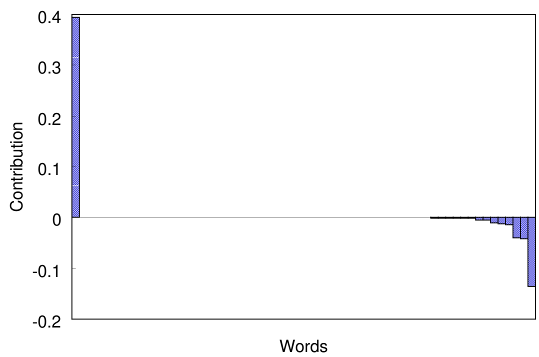


Figure 2: Word contribution between Topic 313 and FT932-6259

with highly negative contribution. On the contrary, most words have a contribution near zero, meaning most words do not have a significant influence on query-document similarity.

As obvious from the definition of word contribution, words with highly positive contribution are presumed to be words that co-occur in the query and document. Such words can be considered as informative words of document relevance to the query. On the contrary, words with highly negative contribution which do not occur in the original query can be considered as words which discriminate relevant documents from other non-relevant documents contained in the data collection.

Since the main objective of query expansion is to add words which are effective in distinguishing relevant documents from the data collection, we assumed that words with highly negative contribution are extremely suitable for expanding the original query. Moreover, we presumed that value of word contribution is a measure of the importance the word has for discrimination. Based on this presumption, the application of word contribution values as the weight of the extracted word for query expansion should also be effective.

### 3.3 Query Expansion Method

Based on our arguments in the previous section, we have developed the following query expansion method.

First, the word contribution of all words in the query and the set of documents from which the words for query expansion are extracted from are calculated. If there are $Num$ documents which are included in the document set for query $q$, the relevant document set $D_{qe}$ can be expressed as $D_{qe}(q) = \{d_1, \cdots, d_{Num}\}$. From each document $d_i$, $N$ words with the lowest contribution are extracted.

Next, a score for each extracted word $w$ is calculated by the following formula:

$$Score(w) = wgt \times \sum_{d \in D_{qe}(q)} Cont(w, q, d)$$

where $wgt$ is a parameter with a negative value (since the contribution of extracted words are also negative). Finally, all extracted words and their scores are added to the original query. If any of the extracted words occur in the original query, that word is not added to the new query. Words with negative scores are also excluded from the expanded query.

## 4 Experiments

In this section, we will describe the experiments made to evaluate our query expansion method.

### 4.1 Preliminary Experiments

From the observation of word contribution data, we discovered that words which occur as a result of morphological analysis errors often have a highly negative contribution. Such "words" include terms with numbers, parantheses, or other punctuation marks. Examples of some of these data are: `[propose`, `0.1p`, `ID=JPRS-JST-003C-18A`, etc. These meaningless words must be deleted from the frequency tables of the documents in order to make an effective retrieval and query expansion.

Based on an empirical theory that these words do not occur frequently, words which occur less than a minimal number in all of the documents were excluded when calculating similarities. As a preliminary experiment, we set several minimal occurence thresholds, and made an evaluation of the text retrieval based on each threshold, by executing a search on TREC-6 data (Topics 301-350). The average precision, R-precision, and the number of retrieved relevant documents for each threshold are shown in Table 2.

Table 2: Retrieval results for baseline search on TREC-6 data

| *min* | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| 0 | 0.0388 | 0.0730 | 425 |
| 1 | 0.0418 | 0.0752 | 425 |
| 2 | 0.0423 | 0.0747 | 429 |
| 4 | 0.0394 | 0.0746 | 439 |
| 8 | 0.0439 | 0.0781 | 471 |
| 16 | 0.0442 | 0.0772 | 488 |
| 24 | 0.0452 | 0.0832 | 502 |
| 32 | 0.0435 | 0.0815 | 530 |
| 48 | 0.0449 | 0.0825 | 573 |
| 64 | 0.0459 | 0.0818 | 599 |

As apparent from the results in Table 2, there was not much difference between the results of these experiments. Therefore, considering the reasonability of threshold values, we decided to set the minimum threshold to 16 and 32 for our following experiments. Therefore, the results for $min = 16, 32$ are used as the baseline for our evaluations on TREC-6 data. The baseline for TREC-7, i.e., TREC-7 retrievals when $min = 16$ and 32, are shown in Table 3.

Table 3: Retrieval results for baseline search on TREC-7 data

| $min$ | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| 16 | 0.0442 | 0.0660 | 448 |
| 32 | 0.0435 | 0.0680 | 470 |

## 4.2 Query Expansion with Relevance Feedback

In order to examine the effectiveness of our query expansion method, we first made query expansion based on relevance feedback.

As described in previous sections, there are 4 parameters for our query expansion method: the minimum word occurence threshold ($min$), the number of documents for query expansion ($Num$), the number of words extracted from each document ($N$), and the weight applied to each extracted word and its contribution ($wgt$). For the description of the experiment results, we will use the following format:

$$\text{rel}.min.Num.N.wgt$$

For example, if $min = 16$, $Num = 10$, $N = 10$, and $wgt = -50$, the run for such conditions will be written as "rel.16.10.10.50".

The document set from which words for query expansion are extracted from is selected based on the results of the baseline search. Of all relevant documents for each query, the top $Num$ ranked documents were extracted. If there were less than $Num$ relevant documents for a query, then all relevant documents were included in the document set.

These experiments were made on TREC-6 and TREC-7 data. Results on TREC-6 data are described in Table 4.

As apparent from these results, we have achieved a significant improvement in both precision and recall compared to the baseline results. The influence of $wgt$ was rather clear: the recall increases and the precision decreases as the absolute value of $wgt$ increases.

Table 4: Retrieval results for query expansion with relevance feedback on TREC-6 data

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| rel.16.10.10.20 | 0.0877 | 0.1484 | 795 |
| rel.16.10.10.50 | 0.0818 | 0.1414 | 954 |
| rel.32.10.10.20 | 0.0932 | 0.1493 | 866 |
| rel.32.10.10.50 | 0.0839 | 0.1422 | 992 |

For further analysis, the precision-recall curveline for the baseline and the query expansion results are presented in Figures 3 and 4.

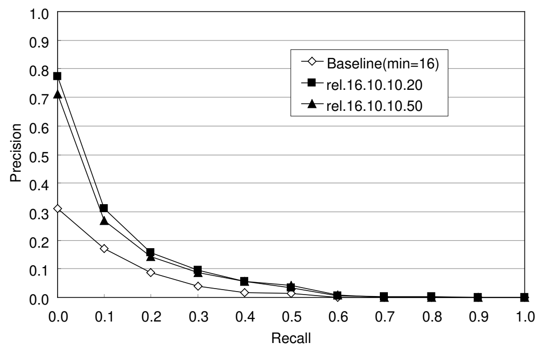


Figure 3: Precision-recall curveline for TREC-6 data ($min = 16$)

The results illustrated in these Figures also prove the effectiveness of our query expansion method.

Next, we will present the results of this experiment made on TREC-7 data. The average precision, R-precision, and number of retrieved relevant documents are shown in Table 5.

Table 5: Retrieval results for query expansion with relevance feedback on TREC-7 data

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| rel.16.10.10.20 | 0.0541 | 0.1134 | 652 |
| rel.16.10.10.50 | 0.0551 | 0.1171 | 848 |
| rel.32.10.10.20 | 0.0510 | 0.1154 | 740 |
| rel.32.10.10.50 | 0.0513 | 0.1140 | 902 |

These results also show an improvement from the baseline retrieval, but the improvement is not as high as TREC-6 experiments. We will also present the precision-recall curveline for TREC-7 in Fig-

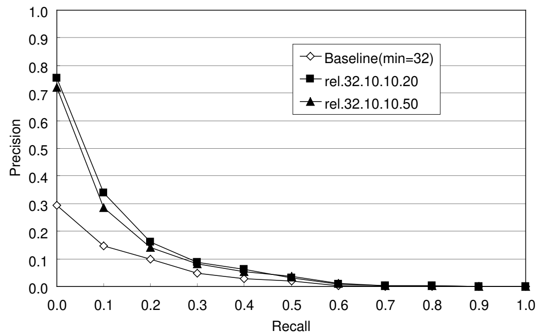


Figure 4: Precision-recall curveline for TREC-6 data ($min = 32$)

ures 5 and 6.

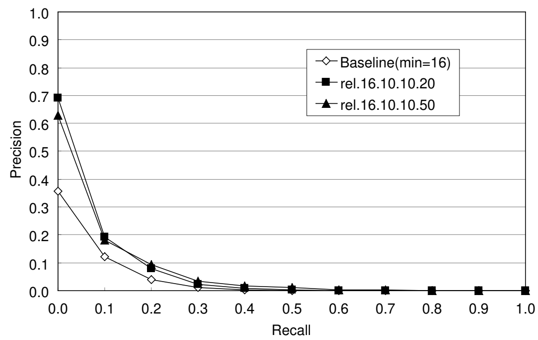


Figure 5: Precision-recall curveline for TREC-7 data ($min = 16$)

Observations from these Figures also show the lack of improvement compared to TREC-6 experiments. The most notable characteristic of TREC-7 results are the drastic descent of precision as the recall increases. Furthermore, the infuence of $wgt$ is different from the results of TREC-6 experiments. As clear from Table 5, the increase of the absolute value of $wgt$ results in the improvement of both the recall *and* precision.

However, overall results of the experiments described in this section proved the effectiveness of the query expansion method based on word contribution with relevance feedback.

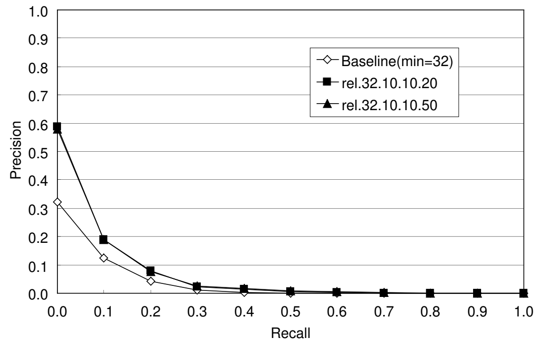


Figure 6: Precision-recall curveline for TREC-7 data ($min = 32$)

## 4.3 Query Expansion with Pseudo Feedback

Since the relevance of documents to a query is unknown in practical use of text retrieval systems, it is essential to develop an effective algorithm of retrieval without using relevance feedback information. Therefore, many systems using the idea of a pseudo feedback, i.e., selecting the top ranked documents from the pilot search as the document set used for query expansion, have been presented in recent years [2][3]. We have made experiments to apply pseudo feedback to our query expansion method. We will describe these experiments in this section.

The difference between the query expansion method described in the previous section based on relevance feedback and the query expansion method for this experiment is the extraction of the $Num$ documents used for word extraction. In this experiment, we extracted the top $Num$ documents of the baseline search as the set of documents used for query expansion, regardless of their actual relevance to the query.

We will use a format similar to the format used in the previous section for the expression of experiment conditions:

$$\text{pse}.min.Num.N.wgt$$

Similar to the previous section, we made experiments on both TREC-6 and TREC-7 data. Detailed experiments were made on TREC-6 data for analysis of the effects of various parameters, and a few experiments were made on TREC-7 data to confirm results. In Table 6, the results for all experiments on TREC-6 data are shown.

Table 6: Retrieval results for query expansion with pseudo feedback on TREC-6 data

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| pse.16.10.10.20 | 0.0161 | 0.0387 | 416 |
| pse.16.10.10.50 | 0.0137 | 0.0300 | 519 |
| pse.16.10.10.100 | 0.0121 | 0.0232 | 518 |
| pse.16.10.20.50 | 0.0140 | 0.0310 | 524 |
| pse.32.5.10.20 | 0.0253 | 0.0543 | 445 |
| pse.32.5.10.50 | 0.0142 | 0.0314 | 467 |
| pse.32.10.5.20 | 0.0191 | 0.0191 | 468 |
| pse.32.10.10.20 | 0.0198 | 0.0402 | 482 |
| pse.32.10.10.50 | 0.0150 | 0.0280 | 547 |
| pse.32.10.10.100 | 0.0131 | 0.0246 | 533 |
| pse.32.10.20.50 | 0.0154 | 0.0280 | 558 |

As obvious from these results, our query expansion method did not improve the retrieval compared to the baseline search. The fact that the experiments with relatively high results used queries in which the expanded words had little influence, also back up the failure of our query expansion.

We will also present the results for the pseudo feedback experiment on TREC-7 data on Table 7. The conditions which had relatively good results from the experiments on TREC-6 data were selected for these experiments.

Table 7: Retrieval results for query expansion with pseudo feedback on TREC-7 data

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| pse.16.10.10.20 | 0.0125 | 0.0373 | 503 |
| pse.16.10.10.50 | 0.0104 | 0.0285 | 594 |
| pse.16.10.10.100 | 0.0085 | 0.0216 | 583 |
| pse.32.10.10.20 | 0.0153 | 0.0387 | 536 |
| pse.32.10.10.50 | 0.0101 | 0.0265 | 570 |
| pse.32.10.10.100 | 0.0085 | 0.0216 | 583 |

As apparent from the data on Table 7, the results for experiments on TREC-7 were no better than those for the TREC-6 experiments. From these results, we presume that there are problems in our query expansion method using pseudo feedback.

# 5 Discussion

In this section, we will discuss the results of our evaluation experiments, and investigate the causes for the failure of our query expansion method.

## 5.1 Analysis of query expansion with relevance feedback

Although we have achieved significant improvement on our query expansion experiment using relevance feedback, consideration is necessary for improvement. As observed from the precision-recall curves illustrated in Figures 3-6, the precision of the retrieval descends rapidly as the recall increases. We examined the word contribution data used for query expansion in these experiments to investigate the cause of this phenomenon.

In Section 3, we explained that there are only a small number of words which have highly negative contribution. Further analysis of word contribution to the similarity between queries and relevant documents showed that, in many cases, there are 1 or 2 words per query-document set that have an extremely high absolute value of word contribution. An example of such data is illustrated in Figure 7.

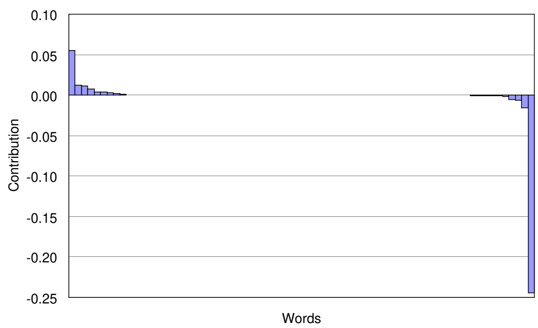


Figure 7: Word contribution between Topic 301 and FBIS3-10397

Since the values of $wgt$ in our experiments were set so that the weight of extracted words would not be extremely higher than the frequencies of words in the original query, there is a strong possibility that the words other than the words with extremely high contribution did not apply sufficient influence on the expanded query.

For the investigation of this hypothesis, we ran query expansion experiments with $wgt = 1200$, so that the other words will have similar weights as the original frequency table. The results are shown in Table 8.

Table 8: Retrieval results for query expansion with relevance feedback ($wgt$ = 1200)

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| (*TREC-6*) rel.16.10.10.1200 | 0.0916 | 0.1454 | 1309 |
| (*TREC-7*) rel.16.10.10.1200 | 0.0688 | 0.1390 | 1336 |

From the comparison of these results and the results presented in Section 4, it is clear that the drastic increasement of $wgt$ has improved both the recall and precision of the retrieval. For comparison, we present the precision-recall curveline for the retrieval on TREC-7 data with $wgt$ as 20 and 1200. This is illustrated in Figure 8.

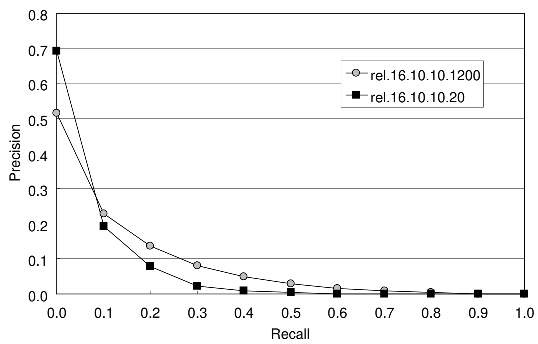


Figure 8: Precision-recall curveline for TREC-7 data ($wgt = 20, 1200$)

As observed from Figure 8, the precision at low recall of rel.16.10.10.1200 is not as good as rel.16.10.10.20. However, the precision of rel.16.10.10.1200 at higher recall is constantly higher than that of rel.16.10.10.20. From these results, we presume that the words with extremely high negative contribution are effective for retrieving documents at a low recall, while the other words extracted for query expansion are effective for retrieving a wide range of relevant documents. Therefore, it is necessary to merge these weights effectively in order to apply the characteristics of each set of words. A reduction or normalization of the extremely high contribution values, such as adapting a logarithm to the word contribution value, may be effective. We have yet to evaluate such methods.

## 5.2 Analysis on query expansion with pseudo feedback

One obvious cause of the failure of our query expansion method with pseudo feedback is the poor precision of the baseline search. Experiments were made on the TREC-6 data with the parameter $Num$ set at 5, 10, and 20. The precision of the baseline retrievals at documents 5, 10, and 20 are shown in Table 9.

Table 9: Precision at 5,10,20 documents for baseline searches

| | TREC-6 | | TREC-7 | |
|---|---|---|---|---|
| | $min$=16 | $min$=32 | $min$=16 | $min$=32 |
| @ 5 | 0.1520 | 0.1560 | 0.1680 | 0.1480 |
| @ 10 | 0.1480 | 0.1520 | 0.1520 | 0.1520 |
| @ 20 | 0.1170 | 0.1160 | 0.1320 | 0.1200 |

As apparent from these results, 83%-88% of the documents used for query expansion were actually irrelevant to the query. This should have a negative effect on the results of query expansion.

In order to examine the effects of a poor baseline search, we simulated the TF∗IDF based retrieval algorithm and the Rocchio feedback based query expansion method applied in the SMART system at TREC-7[3]. The Rocchio weights were calculated by the following formula:

$$\vec{Q_{new}} = \alpha \times \vec{Q_{org}} + \beta \times \frac{1}{R} \sum_{D \in Rel} \vec{D} - \gamma \times \frac{1}{N} \sum_{D \notin Rel} \vec{D}$$

where $\alpha = 3$, $\beta = 2$, $\gamma = 2$, and 20 new terms with the highest Rocchio weights for each query were added to the original query. These parameter values were set as the values presented in the SMART paper by AT&T on TREC-7. However, SMART also added 5 new phrases in this process. Since we do not have any indexing methods especially tuned for phrases, this function was not applied in our simulation.

The average precision, R-precision and number of retrieved relevant documents by the baseline search of SMART and the Rocchio feedback based query expansion are shown in Table 10, and the precision-recall curveline for these SMART retrievals are illustrated in Figure 9.

Table 10: Retrieval results for query expansion with pseudo feedback on TREC-7 data

| Condition | Avg Prec | R-Prec | Rel-ret |
|---|---|---|---|
| Baseline | 0.1433 | 0.1848 | 1887 |
| Rocchio | 0.1348 | 0.1691 | 1392 |

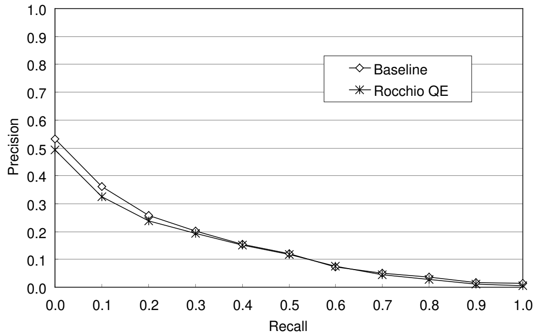


Figure 9: Precision-recall curveline for SMART algorithm

As apparent from Table 10, the result of the search by the SMART algorithm was good, but it is not as good as the results presented in the TREC-7 papers (AT&T's average precision was 0.2290.).

Furthermore, the results from Table 10 and Figure 9 show that the Rocchio feedback based query expansion could not exceed the baseline retrieval, which was similar to the case in our experiments. These results prove that a poor baseline retrieval has a negative impact on the results of query expansion, even when applying an established query expansion algorithm. Therefore, the lack of precision of the baseline search can be considered as a cause of our poor results.

However, further analysis on our data showed that queries with many relevant documents included in the document set for query expansion also had poor results after expanding the query. Table 11 is a list of topics of which the precision of the baseline retrieval at 10 documents were higher than or equal to 0.60, and the average precision of the baseline ($min = 16$) and query expansion (qe.16.10.10.50).

As obvious from these results, the high ratio of relevant documents included in the document set for query expansion did not improve the results of query expansion.

Table 11: Retrieval results for topics with high precision at 10 documents

| Topic | Prec @ 10 | Baseline | qe.16.10.10.50 |
|---|---|---|---|
| 302 | 0.90 | 0.2113 | 0.1286 |
| 314 | 0.70 | 0.2074 | 0.0468 |
| 353 | 0.60 | 0.0778 | 0.0823 |
| 357 | 0.80 | 0.1054 | 0.0361 |
| 368 | 0.60 | 0.0752 | 0.0508 |
| 398 | 1.00 | 0.1482 | 0.0095 |

The main idea behind our query expansion method may be an explanation of this result. As described in Section 3, the extraction of words with highly negative contribution was based on the hypothesis that such words are discriminant of the concerned document. If these words were extracted from documents which were not relevant to the original query, the expanded query will contain highly discriminant words of non-relevant documents. It is quite obvious that such query expansion will decrease the precision of retrieval.

Another cause may be the weighting problems which were pointed out in the previous analysis on relevance feedback experiments. In many cases, 1 or 2 extracted words have an extremely high weight after query expansion, as previously explained. This means that the discriminant words extracted from non-relevant documents will be extremely high weighted after query expansion. Therefore, the mere existence of non-relevant documents in the document set for query expansion can make a large negative influence on the final retrieval results.

However, it is difficult to make a strict failure analysis on the query expansion method if the indexing is erroneous. We suspect that this is the main cause of our poor results, since the text retrieval algorithm of SMART also did not achieve satisfying results with our frequency tables. Bugs on our dictionary are especially crucial with our current method, since words extracted for query expansion are observed to be words with relatively low term and document frequency, which may result from such bugs. Considering the fact that the contributions of the extracted words seem to be sensitive to the scarcity of the word, we believe that the improvement of our morphological analysis program is essential for strict evaluation.

# 6 Conclusion

In this paper, we have proposed a novel query expansion method based on word contribution, which is a measure of the influence of a word to query-document similarity. Through the analysis of word contribution on queries and relevant documents, we set a hypothesis that words with highly negative contribution are words which discriminate relevant documents from other documents in the data collection. Based on this hypothesis, we developed a query expansion method which adds such words and their weighted contribution to the original query.

First, we evaluated our query expansion by experiments using relevance feedback information. Results from these experiments proved the effectiveness of our proposed method. Second, we made evaluation experiments based on pseudo feedback. The results from these experiments were dissatisfying. Through the analysis of our results, we came to the conclusion that an improvement on the weighting of extracted words was necessary.

However, simulation of an established query expansion method on our data showed that an improvement on the indexing process, or, in other words, the dictionary used for our morphological analysis program was also necessary for the improvement of our results. We believe improvement of the morphological analysis program (or the application of a common-used program) is indispensable for future development.

One of our future studies will be the improvement of the weighting formula for extracted words. We consider it necessary to develop a new weighting method to cope with the words with extremely high contribution values. We also want to examine the word contribution of query-document similarity of non-relevant documents, which we have not made detailed analysis yet. Analysis on non-relevant documents should be helpful in our relevance feedback method.


# Acknowledgments

The authors would like to appreciate the researchers in the Knowledge-Based Information Processing Lab of KDD R&D Laboratories for their fruitful opinions. We will also express gratitude to Shigeki Ohira of Waseda University and Marko Herzog of HTW Dresden for their tremendous efforts on our participation to TREC.


# References


[1] M Iwayama, and T Tokunaga: "A Probabilistic Model for Text Categorization: Based on a Single Random Variable with Multiple Values", Proceedings of 4th Conference on Applied Natural Language Processing, pp.162-167, 1994.

[2] S Robertson, S Walker, S Jones, M Hancock-Beaulieu, and M Gatford, "Okapi at TREC-3", Overview of the Third Text REtrieval Conference, pp 109-125, 1994.

[3] A Singhal, J Choi, D Hindle, D Lewis, and F Pereira: "AT&T at TREC-7", The Seventh Text REtrieval Conference, 1998. (to be published)