Multilingual Relevant Sentence Detection Using Reference Corpus

Ming-Hung Hsu, Ming-Feng Tsai, and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan
{mhhsu, mftsai}@nlg.csie.ntu.edu.tw
hhchen@csie.ntu.edu.tw

Abstract. IR with reference corpus is one approach when dealing with relevant sentences detection, which takes the result of IR as the representation of query (sentence). Lack of information and language difference are two major issues in relevant detection among multilingual sentences. This paper refers to a parallel corpus for information expansion and translation, and introduces different representations, i.e. sentence-vector, document-vector and term-vector. Both sentence-aligned and document-aligned corpora, i.e., Sinorama corpus and HKSAR corpus, are used. The factors of aligning granularity, the corpus domain, the corpus size, the language basis, and the term selection strategy are addressed. The experiment results show that MRR 0.839 is achieved for similarity computation between multilingual sentences when larger finer grain parallel corpus of the same domain as test data is adopted. Generally speaking, the sentence-vector approach is superior to the term-vector approach when sentence-aligned corpus is employed. The document-vector approach is better than the term-vector approach if document-aligned corpus is used. Considering the language issue, Chinese basis is more suitable to English basis in our experiments. We also employ the translated TREC novelty test bed to evaluate the overall performance. The experimental results show that multilingual relevance detection has 80% of the performance of monolingual relevance detection. That indicates the feasibility of IR with reference corpus approach in relevant sentence detection.

1 Introduction

Relevance detection on sentence level aims to identify relevant sentences from a collection of sentence set given a specific topic specification. Since it is an elementary task in some emerging applications like multi-document summarization and question-answering, it has attracted many researchers' attentions recently. The challenging issue behind sentence relevance detection is: the surface information that can be employed to detect relevance is much fewer than that in document relevance detection. TREC (Harman, 2002) organized a relevance/novelty detection track starting from 2002 to evaluate the technological development of this challenging problem.

S. H. Myaeng et al. (Eds.): AIRS 2004, LNCS 3411, pp. 165 – 177, 2005.
© Springer-Verlag Berlin Heidelberg 2005

In the past, several approaches were proposed to identify sentence relevancy. Word matching and thesaurus expansion were adopted to recognize if two sentences touched on the same subject in multi-document summarization (Chen, *et al.*, 2003). Such an approach has been employed to detect relevance between a topic description and a sentence (Tsai and Chen, 2002). Zhang *et al.* (2002) employed an Okapi system to retrieve relevant sentences with queries formed by topic descriptions. Allan *et al.* (2003) focused on the novelty detection algorithms and showed how the performance of relevant detection affects that of novelty detection. Instead of using an IR system to detect relevance of sentences directly, a reference corpus approach has been proposed (Chen, Tsai and Hsu, 2004). In this approach, a sentence is considered as a query to a reference corpus, and two sentences are regarded as similar if they are related to the similar document lists returned by IR systems.

The above approaches focus on monolingual relevance sentence detection only. As all know, large scale multilingual data have been disseminated very quickly via Internet. How to extend the applications to multilingual information access is very important. Chen, Kuo and Su (2003) touched on multilingual multidocument summarization. To measure the similarities between two bilingual sentences is their major concern.

This paper extends the reference corpus approach (Chen, Tsai and Hsu, 2004) to identify relevant sentences in different languages. The computation of the similarities between an English sentence and a Chinese sentence, which is the kernel of multilingual relevant sentence detection, will be studied by referencing sentence-aligned and document-aligned parallel corpora. Section 2 introduces the basic concepts of the reference corpus approach, and a multilingual Okapi-based IR system used in our experiments. Section 3 shows its extension to multilingual relevance detection. Section 4 presents our reference corpora and evaluation criteria. Section 5 shows and discusses the experimental results. Section 6 further compares the performance differences between monolingual and multilingual relevance detection on TREC evaluation data. Section 7 concludes the remarks.

2 Relevance Detection Using Reference Corpus

To use a similarity function to measure if a sentence is on topic is similar to the function of an IR system. We use a reference corpus, and regard a topic and a sentence as queries to the reference corpus. An IR system retrieves documents from the reference corpus for these two queries. Each retrieved document is assigned a relevant weight by the IR system. In this way, a topic and a sentence can be in terms of two weighting document vectors. Cosine function measures their similarity, and the sentence with similarity score larger than a threshold is selected. The issues behind the IR with reference corpus approach include the reference corpus, the performance of an IR system, the number of documents consulted, the similarity threshold, and the number of relevant sentences extracted.

The reference corpus should be large enough to cover different themes for references. Chen, Tsai, and Hsu (2004) consider TREC-6 text collection as a reference corpus. Two IR systems, i.e., Smart and Okapi, were adopted to measure the effects of the performance of an IR system. Their experimental results show that Okapi-based relevance detector outperforms Smart-based one. Thus Okapi system is adopted in the latter experiments.

We modify Okapi-Pack[1] from City University (London) to support Chinese information retrieval in the following way. A Chinese word-segmentation system is used for finding word boundaries. Unknown words may be segmented into a sequence of single Chinese characters. While indexing, Okapi will merge continuous single characters into a word and treat it as an index term. We build a single-character word list to avoid merging a single-character word into an unknown word. Chinese stop word list is not adopted.

We adopted NTCIR3 Chinese test collection (Chen, *et al.*, 2003) to evaluate the performance of Chinese Okapi system (called *C-Okapi* hereafter). Table 1 summarizes the performance of C-Okapi comparing to the results of the participants in NTCIR3 (Chen, *et al.*, 2003). The first column denotes different query construction methods, where T, C, D, and N denote topic, concept, description, and narrative, respectively. The 2^{nd}-4^{th} columns, i.e., AVG, MAX, and MIN, denote the average, the maximum, and the minimum performance, respectively. C-Okapi outperforms or competes with the maximum one in T and C methods, and is above the average in the other two query construction methods. In the later experiments, we will adopt Okapi and C-Okapi for bilingual relevance detection.

Table 1. Performance of C-Okapi

Topic Field	AVG	MAX	MIN	C-Okapi
C	0.2605	0.2929	0.2403	0.2822
T	0.2467	0.2467	0.2467	0.2777
TC	0.3109	0.3780	0.2389	0.3138
TDNC	0.3161	0.4165	0.0862	0.3160

3 Similarity Computation Between Multilingual Sentences

In Section 2, we consult a monolingual corpus to determine the similarity between two sentences in the same language. When this approach is extended to deal with multilingual relevance detection, a parallel corpus is used instead. This corpus may be document-aligned or sentence-aligned. Figure 1 shows the overall procedure. English and Chinese sentences, which are regarded as queries to a parallel corpus, are sent to Okapi and C-Okapi, respectively. Total *R* English and Chinese documents/sentences[2] accompanying with the relevance weights are retrieved for English and Chinese que-

[1] http://www.soi.city.ac.uk/~andym/OKAPI-PACK/

[2] The word documents/sentences denotes either document-aligned or sentence-aligned corpus is used.

ries. Because the corpus is aligned, the returned document (or sentence) IDs are comparable. Cosine function is used to compute the similarity, and thus the degree of relevance.

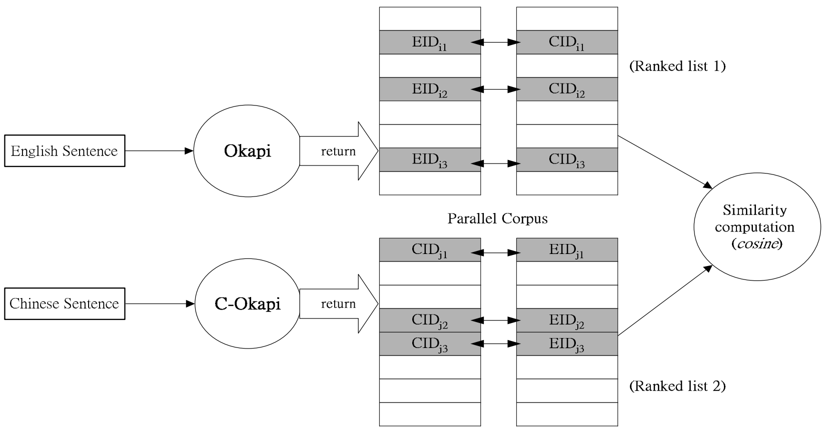


Fig. 1. Document-Vector/Sentence-Vector Approach

In the above, two sentences are considered as relevant if they have similar behaviors on the results returned by IR systems. The results may be ranked list of documents or sentences depending on the aligning granularity of the parallel corpus. Besides the *document-vector/sentence-vector approach* shown in Figure 1, the two vectors used in similarity computation may be in terms of relevant terms. This idea follows the corpus-based approach to query translation (Davis and Dunning, 1995) in cross language information retrieval (CLIR). In CLIR, a query in language *A* is submitted to an *A*-*B* parallel corpus. An IR system for language *A* selects the relevant documents in *A*. The documents in language *B* are also reported at the same time. The target query is composed of terms selected from the relevant documents in *B*, and finally submitted to IR system for *B* language.

The above procedure is considered as *translation* in CLIR. Now, the idea is extended and plays the roles of both *translation* and *information expansion*. Figure 2 shows the overall flow. Similarly, an English sentence and a Chinese sentence, which will be determined relevancy, are sent to the two IR systems. *R* most relevant documents/sentences in two languages are returned. Instead of using the retrieval results directly, we select *K* most representative terms from the resulting documents/sentences. The two sets of *K* terms form two vectors, so that this approach is called *term-vector approach* later. Cosine function determines the degree of relevance between the English and the Chinese sentences.

Because the *R* most relevant documents/sentences are in two languages, we can consider either English or Chinese documents/sentences as a basis. In other words, if we select Chinese, then we map ranked list 1 (i.e., English results) into Chinese correspondent through the document-aligned/sentence-aligned chains. Similarly, ranked list 2 (i.e., Chinese results) may be mapped into English correspondent when English part is selected as a basis. Now we consider how to select the *K* most representative terms. Two alternatives shown below are adopted.

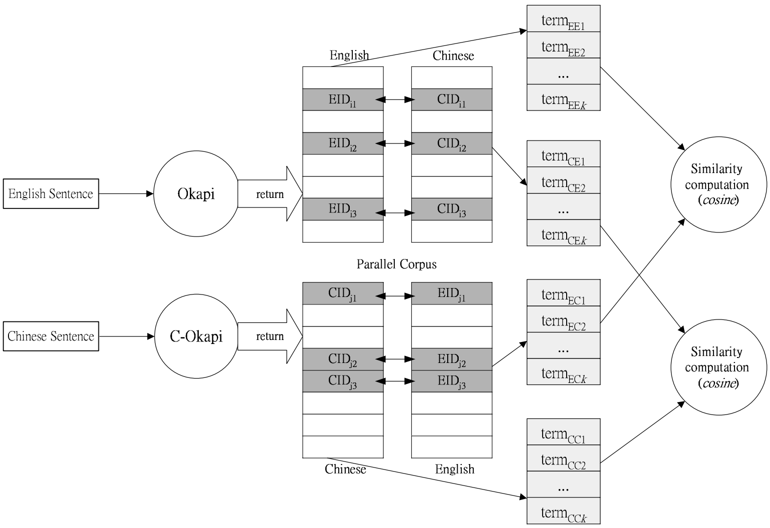


Fig. 2. Term-Vector Approach

3.1 Weighting Scheme: Okapi-FN1

An intuitive weighting scheme is the weighting function of IR system. The weighting function of a term *t* in Okapi is as follows:

$$W(t) = \log \frac{(r+0.5)(N-R-n+r+0.5)}{(R-r+0.5)(n-r+0.5)} \ . \tag{1}$$

where *N* is total number of documents/sentences in the reference parallel corpus, *R* is the number of relevant documents/sentences to a query, *n* is the number of documents/sentences in which term *t* occurs, and *r* is the number of relevant documents/sentences in which term *t* occurs.

In our experiments, top *R* (different values of *R* are tested) documents in the ranked list are parsed and the total occurrences *r* of a term *t* in the *R* documents are counted. Terms with the top *K* weights are employed for similarity computation.

3.2 Weighting Scheme: Log-Chi-Square

The *Chi-Square* test is used to find the terms highly relevant to the returned documents/sentences. Besides, Chi-Square test is also considered as a basis for weighting. A 2×2 contingency table shown in Table 2 is conducted for Chi-Square test.

Table 2. Contingency Table for Chi-Square Test

	Relevant documents/sentences	Non-relevant documents/sentences
Term *t* occurs	A = *r*	B = *n* – *r*
Term *t* not occur	C = R – r	D = N – R – (n – r)

The meanings of N, n, R, and r are the same as those described in Okapi-FN1. The formula for Chi-Square test is shown as follows:

$$\text{Chi-Square test } \chi^2 = \frac{N(AD-BC)^2}{(A+B)(A+C)(B+D)(C+D)} . \tag{2}$$

For the value of χ^2 could be very large (even larger than 10^6), we take logarithm of χ^2 as the weight of a term to avoid the cosine value between two vectors to be dominated by some few terms. This operation is similar to smoothing and drops the scale of weights.

Summing up, two alternatives, i.e., vectors in terms of resulting documents/sen tences (Figure 1), and vectors in terms of representative terms (Figure 2), may be considered for similarity computation in multilingual relevance detection. In the latter case, either English part or Chinese part may be considered as a basis, and each has two possible weight schemes, i.e., Okapi-FN1 and Log-Chi-Square. Thus, four possible combinations are conducted in total for the latter experiments.

4 Experiment Materials and Evaluation Method

Two Chinese-English aligned corpora are referenced in our experiments. One is Sinorama corpus[3], and the other one is HKSAR Corpus[4]. Sinorama consists of documents published by Sinorama magazine within 1976-2001. This magazine, which is famous for her superior Chinese-English contrast, recorded Taiwan society's various dimensions of evolvements and changes. HKSAR collects news articles released by the Information Services Department of Hong Kong Special Administrative Region (HKSAR) of the People's Republic of China. The following compares these two corpora from corpus scale, aligning granularity, average length, and so on.

Sinorama is a "sentence-aligned" parallel corpus, consisting of 50,249 pairs of Chinese and English sentences. We randomly select 500 Chinese-English pairs as test data to simulate multilingual relevance sentence detection. The remaining 49,749 pairs are considered as a parallel reference corpus. They are indexed separately as two monolingual databases, in which a Chinese sentence or an English sentence is regarded

[3] http://rocling.iis.sinica.edu.tw/ROCLING/corpus98/mag.htm

[4] http://wave.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2000T46

as a "small document". The average length of Chinese sentences in the reference corpus is 151 bytes and that of English sentences is 254 bytes. The average length of Chinese and English test sentences is 146 and 251 bytes, respectively.

HKSAR corpus contains 18,147 pairs of aligned Chinese-English documents released by HKSAR from July 1, 1997 to April 30th, 2000. Similarly, we index all articles in the same language as a monolingual database. The average document length is 1,570 bytes in Chinese and 2,193 bytes in English. The test data used in experiments are the same sentences pairs as Sinorama.

At first, we develop an evaluation method and a set of experiments to measure the kernel operation of relevance detection only, i.e., the similarity computation between Chinese and English sentences, in Section 5. Then, we measure the overall performance of multilingual relevance detection in Section 6. As mentioned, total 500 pairs of Chinese-English sentences are randomly selected from Sinorama corpus. They are denoted as: $<C_1, E_1>$, $<C_2, E_2>$, …, $<C_{500}, E_{500}>$, where C_i and E_i stand for Chinese and English sentences, respectively. Among the 500 Chinese sentences C_1, …, C_{500}, C_i is the most relevant to E_i. In other words, when we compute the similarities of all combinations consisting of one Chinese and one English sentences, C_i should be the most similar to E_i for $1 \leqq i \leqq 500$ ideally. Let $Sim(i, j)$ be the similarity function between C_i and E_j. A match function $RM(i, j)$ is defined as follows:

$$RM(i, j) = |\{k | \, Sim(i, k) > Sim(i, j), \; 1 \leqq k \leqq 500\}| + 1 \; . \tag{3}$$

The match function assigns a rank to each combination. The perfect case is $RM(i, i)=1$. We call it a *perfect match* later. We also relax the case. If $RM(i, i)$ is no larger than a threshold, we consider the result of matching is "good". In our experiments, the threshold is set to 10. That is, we postulate that the first 2% of matching pairs will cover the correct matching.

Consulting the evaluation method in question answering track of TREC, we adopt **MRR** (mean reciprocal rank) score to measure the performance. Let $S(i)$ be the evaluation score for a topic i (Chinese sentence). MRR is summation of $S(i)$.

$$S(i) = \begin{cases} 1 / M(i, i) & \text{if } RM(i, i) \leqq 10 \\ 0 & \text{else} \end{cases} \; . \tag{4}$$

$$\text{MRR} = \frac{1}{500} \sum_{i=1}^{500} S(i) \; . \tag{5}$$

5 Result Discussion

5.1 Using Sinorama Corpus

Sentence Vector Approach. Table 3 shows the experimental result of sentence-vector approach along with Sinorama corpus. Row "$RM(i, i)=1$" denotes how many topics get a "perfect match" and row "$RM(i, i) \leqq 5$" denotes how manytopics get a correct match in the first 5 ranks. For example, 77.40% of test data are perfect match if 200 sentences are consulted by Okapi and C-Okapi, i.e., 200 sentences are returned for reference. In this case, the MRR is 0.839, which is the best in this experiment. When the number of returned sentences increases from 50 to 200, MRR score also increases. Then MRR score goes down until the number of returned sentences reaches about 600. After that, MRR score rises again and reaches to a stable state, i.e., 0.82-0.83. Figure 3 captures the performance change.

Analyzing the result, we find there may be two degrees of relevancy of small documents (i.e., sentences) in the corpus to a query. Documents with high relevance are easily retrieved with ranks smaller than 200. When the rank increases larger than 200, lowly relevant documents are retrieved with more non-relevant documents. That introduces noise for similarity computation. The influence reaches to the worst between ranks 500 and 600, and then goes down since the weights of vector elements are decreased. The other reason may be that some returned sentences in both vectors are complementary when smaller number of sentences is consulted, and the complementary parts show up when more sentences are consulted.

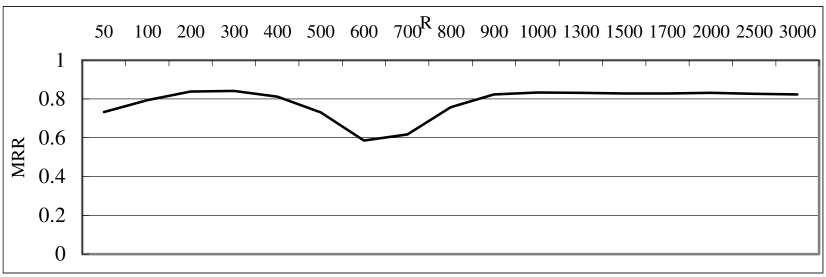


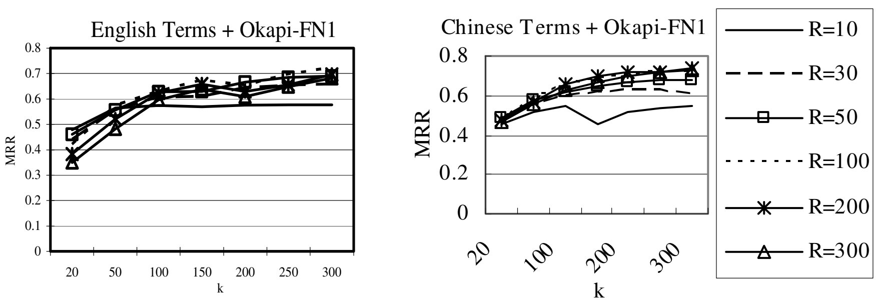
Fig. 3. MRR Score versus Number of Returned Sentences

Term-Vector Approach. Figures 4, 5, 6 and 7 show the results of term-vector approach, where terms in either English or Chinese are used, and two weighting schemes, i.e., Okapi-FN1 and Log-Chi-Square, are applied. The *x* axis represents *k*, the number of terms used for similarity computation. The *y* axis denotes the MRR score.

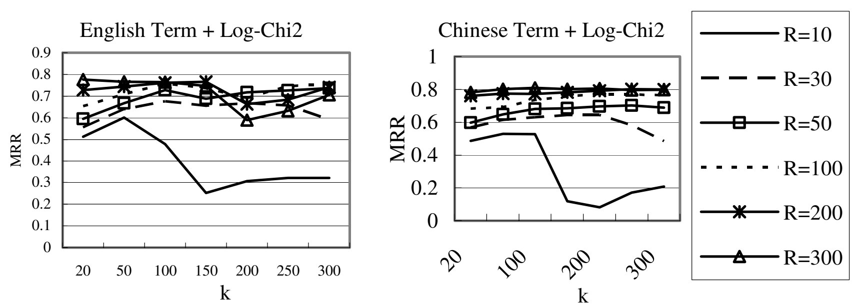
Several interesting conclusions can be made after the factors of language and weighting schemes are considered. Performances of Figures 4 and 5 are inferior to those of Figures 6 and 7. It shows that Log-Chi-Square weighting scheme is more suitable for term-vector approach than Okapi-FN1 weighting scheme. It meets our expectation that Log-Chi-Square weighting scheme properly captures concepts embedded in resulting sentences returned by Okapi and C-Okapi. Performances of the runs of smaller *k* (=50) in the four figures show that Log-Chi-Square scheme will assign higher weights to terms which are truly relevant to the sentence (query).

Observing the differences between Figures 4 and 5, and between Figures 6 and 7, we can find that the trends of performances using terms in different languages are dissimilar. Using English terms as vector elements, the performance trend shows undulation as we saw in Figure 2, though the drops are smaller in Figures 4 and 6. On the other hand, performance trend of using Chinese terms as vector elements is monotonously increasing with *k* when *R* is greater than 30. It may indicate that English suffers from more noises, such as word sense ambiguity, than Chinese.

The best performance, near 0.81, appears in the case "*R*=300" of Figure 7, i.e., take Chinese as a basis and Log-Chi-Square formula. It is lower than the best performance 0.84 in sentence-vector approach. The whole performance of term-vector approach is also inferior to sentence-vector approach.



Figs. 4 and 5. English Terms or Chinese Terms plus Okapi-FN1 Weighting Scheme



Figs. 6 and 7. English Terms or Chinese Terms plus Log-Chi-Square Weighting Scheme

5.2 Using HKSAR Corpus

Document-Vector Approach. Figure 8 shows the results of the application of the document-vector approach on HKSAR corpus, which is a document-aligned Chinese-English corpus. The best one has only 30% of the performance shown in Figure 3. The result shows the influence of corpus domain on reference corpus approach. Since the 500 pairs of test sentences are randomly selected from Sinorama corpus, the domain of test sentences and the reference databases are the same, i.e., content focused on major events and construction in Taiwan from 1976-2001. In contrast, the HKSAR corpus contains the news issued by HKSAR within 1997-2000. The test sentences and the reference corpus are totally different in domain of concepts so that there are rarely relevant documents in HKSAR. That introduces much more noises than useful information in ranked list. Besides the domain issue, the small size of HKSAR corpus results in poor performance in retrieval too.

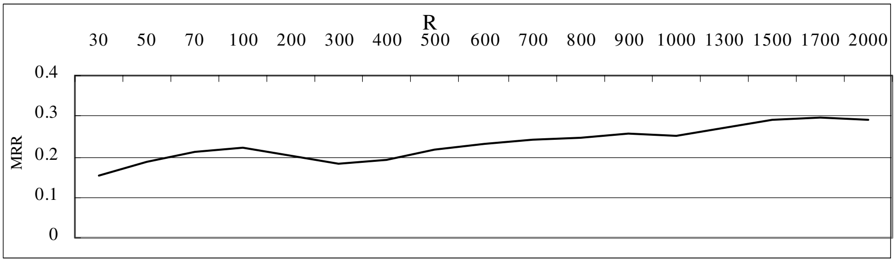
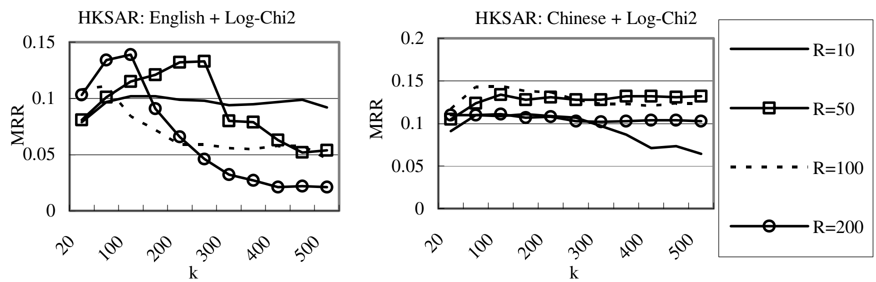


Fig. 8 Document-Vector Approach Using HKSAR

Term-Vector Approach. Figures 9 and 10 show the results of term vector approach on HKSAR, using Log-Chi-Square weighting scheme. Chinese-term-based approach (Figure 10) is more robust than English-term-based approach (Figure 9). However, their performance does not compete with that of document-vector approach. As HKSAR is a "document-aligned" parallel corpus, it is more difficult to select terms suitable for information expansion. Thus, the performance goes down from Figure 8 to Figure 9 and Figure 10. The performance drop is more obvious than that between Figures 3 and 7.



Figs. 9 and 10. English and Chinese Terms plus Log-Chi-Square Weighting with HKSAR

6 Experiments of Multilingual Relevance Detection

Besides evaluating the similarity computation, we also employ the test data in TREC 2002 Novelty track to evaluate the overall multilingual relevance detection. The test data includes 49 topics, each of which is given a set of sentences to evaluate the performance of relevance detector (Harman, 2002). All of these topics and sentences are in English. For multilingual relevant sentence detection, all topics are manually translated into Chinese. Each translated topic (in Chinese) and the corresponding given set of sentences (in English) are sent as queries to C-Okapi and Okapi respectively, so that we can compute similarity between each topic and each sentence in the given set using either document-vector or term-vector approach.

Chen, Tsai, and Hsu (2004) use logarithmic regression to simulate the relationship between total number of the given sentences and number of the relevant sentences, in TREC 2002 Novelty track. We adopt the similar approach. A dynamic percentage of sentences most similar to topic *t* in the given set will be reported as relevant. According to the assessment of TREC 2002 Novelty track, we can compute precision, recall, and F-measure for each topic. Figure 11 shows the performance, i.e., average F-measure of 49 topics, using Sinorama and HKSAR as reference corpora, respectively. Sentence-vector approach (Section 5.1.1) and document-vector approach (Section 5.2.1) are adopted.

Apparently, using Sinorama as a reference corpus outperforms using HKSAR. This result is consistent with the evaluation in similarity computation. Chen, Tsai, and Hsu (2004) used TREC6 text collection, which consists of 556,077 documents, as reference corpus. The best performance using Sinorama for multilingual relevance detection is about 80% of monolingual relevance detection, i.e., 0.212 (Chen, Tsai, and Hsu), and using HKSAR is about 50%. Note that the human performance in monolingual relevance detection is 0.371.

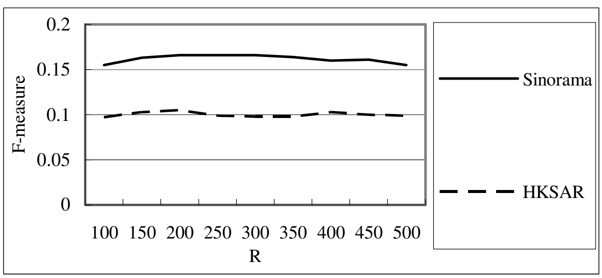


Fig. 11. Sentence/Document-Vector Approach Using TREC 2002 Topics

7 Conclusions and Future Work

This paper considers the kernel operation in multilingual relevant sentence detection. A parallel reference corpus approach is adopted. The issues of aligning granularity, the corpus domain, the corpus size, the language basis, and the term selection strategy are addressed. In the intensive experiments, the best MRR (0.839) is achieved when the test data and the reference corpus come from the same domain, the finer-grained alignment (i.e., sentence alignment), and the larger corpus are adopted. In that case, 77.40% of test data are ranked 1.

Generally speaking, the sentence-vector approach is superior to the term-vector approach when sentence-aligned corpus is employed. The document-vector approach is better than the term-vector approach if document-aligned corpus is used. In term-vector approach, Log-Chi-Square weighting scheme is better than Okapi-FN1 weighting scheme. Considering the language issue, Chinese basis is more suitable to English basis in our experiments. It shows that performance trends may depend on the characteristics of different languages.

Comparing the monolingual and the multilingual relevance detection, the latter has 80% performance of the former. It shows that IR with reference corpus approach is adapted easily to multilingual domain.

From the experiment results, we infer that if the domain of reference corpus is the same as that of query, the performance of relevance detection will be better. While the domain of a query is often unknown, large domain-coverage corpora should be more appropriate than small ones. More, we can infer that the finer-grained alignment corpus is more suitable for multilingual relevant sentence detection. In future work, we'll design more careful experiments to verify the two points and to find out other characteristics of IR with reference corpus approach.

Acknowledgements

Research of this paper was partially supported by National Science Council, Taiwan, under the contracts NSC 93-2213-E-002-078 and NSC 93-2752-E-001-001-PAE.

References

[1] Allan, James; Wade, Courtney and Bolivar, Alvaro (2003) "Retrieval and Novelty Detection at the Sentence Level," *Proceedings of the 26th ACM SIGIR* , 314-321.

[2] Chen, Hsin-Hsi; Kuo, June-Jei; Huang, Sheng-Jie; Lin, Chuan-Jie and Wung, Hung-Chia (2003) "A Summarization System for Chinese News from Multiple Sources," *Journal of American Society for Information Science and Technology*, **54**(13), 1224-1236.

[3] Chen, Hsin-Hsi; Kuo, June-Jei; and Su, Tsei-Chun (2003). "Clustering and Visualization in a Multi-Lingual Multi-Document Summarization System," *Proceedings of 25th European Conference on Information Retrieval Research*, LNCS 2633, 266-280.

[4] Chen, Hsin-Hsi; Tsai, Ming-Feng and Hsu, Ming-Hung (2004) "Identification of Relevant and Novel Sentences Using Reference Corpus," *Proceedings of 26th European Conference on IR Research*, LNCS 2997, 85-98.

[5] Chen, Kuang-hua; Chen, Hsin-Hsi; Kando, Noriko; Kuriyama, Kazuko; Lee, Sukhoon; Myaeng, Sung Hyon; Kishida, Kazuaki; Eguchi, Koji and Kim, Hyeon (2003) "Overview of CLIR Task at the Third NTCIR Workshop," *Proceedings of NTCIR Workshop 3*.

[6] Davis, Mark and Dunning, Ted (1995) "A TREC Evaluation of Query Translation Methods for Multi-lingual Text Retrieval," *Proceedings of TREC 1995*, 483-498.

[7] Harman, Donna (2002) "Overview of the TREC 2002 Novelty Track," *Proceedings of TREC 2002*, NIST SP 500-251, Maryland.

[8] Tsai, Ming-Feng and Chen, Hsin-Hsi (2002) "Some Similarity Computation Methods in Novelty Detection," *Proceedings of TREC 2002*, NIST SP 500-251, Maryland.

[9] Zhang, Min; Song, Ruihua; Lin, Chuan; Ma, Shaoping; Jiang, Zhe; Yijiang Jin; Liu Yiqun; and Zhao, Le (2002) "Expansion-Based Technologies in Finding Relevant and New Information," NIST SP 500-251, Maryland.

[10] http://trec.nist.gov/pubs/trec11/papers/NOVELTY.OVER.pdf

[11] http://trec.nist.gov/pubs/trec11/papers/ntu.feng.final2.pdf.

[12] http://trec.nist.gov/pubs/trec11/papers/tsinghuau.novelty2.pdf