

# Multilingual Spectral Clustering Using Document Similarity Propagation

Dani Yogatama and Kumiko Tanaka-Ishii

Graduate School of Information Science and Technology, University of Tokyo  
13F Akihabara Daibiru, 1-18-13 Kanda Chiyoda-ku, Tokyo, Japan  
yogatama@cl.ci.i.u-tokyo.ac.jp kumiko@i.u-tokyo.ac.jp

## Abstract

We present a novel approach for multilingual document clustering using only comparable corpora to achieve cross-lingual semantic interoperability. The method models document collections as weighted graph, and supervisory information is given as sets of must-linked constraints for documents in different languages. Recursive  $k$ -nearest neighbor similarity propagation is used to exploit the prior knowledge and merge two language spaces. Spectral method is applied to find the best cuts of the graph. Experimental results show that using limited supervisory information, our method achieves promising clustering results. Furthermore, since the method does not need any language dependent information in the process, our algorithm can be applied to languages in various alphabetical systems.

## 1 Introduction

Document clustering is unsupervised classification of text collections into distinct groups of similar documents. It has been used in many information retrieval tasks, including data organization (Siersdorfer and Sizov, 2004), language modeling (Liu and Croft, 2004), and improving performances of text categorization system (Aggarwal et al., 1999). Advance in internet technology has made the task of managing multilingual documents an intriguing research area. The growth of internet leads to the necessity of organizing documents in various languages. There exist thousands of languages, not to mention countless minor ones. Creating document clustering model for each language is simply unfeasible. We need methods to deal with text collections in diverse languages simultaneously.

Multilingual document clustering (MLDC) involves partitioning documents, written in more than one languages, into sets of clusters. Similar documents, even if they are written in different languages, should be grouped together into one cluster. The major challenge of MLDC is achieving cross-lingual semantic interoperability. Most monolingual techniques will not work since documents in different languages are mapped into different spaces. Spectral method such as Latent Semantic Analysis has been commonly applied for MLDC task. However, current techniques strongly rely on the presence of common words between different languages. This method would only work if the languages are highly related, i.e., languages that share the same root. Therefore, we need another method to improve the robustness of MLDC model.

In this paper, we focus on the problem of bridging multilingual space for document clustering. We are given text documents in different languages and asked to group them into clusters such that documents that belong to the same topic are grouped together. Traditional monolingual approach is impracticable since it is unable to predict how similar two multilingual documents are. They have two different spaces which make conventional cosine similarity irrelevant. We try to solve this problem utilizing prior knowledge in the form of must-linked constraints, gathered from comparable corpora. Propagation method is used to guide the language-space merging process. Experimental results show that the approach gives encouraging clustering results.

This paper is organized as follows. In section 2, we review related work. In section 3, we propose our algorithm for multilingual document clustering. The experimental results are shown in section 4. Section 5 concludes with a summary.

## 2 Related Work

Chen and Lin (2000) proposed methods to cluster multilingual documents using translation technology, relying on cross-lingual dictionary and machine-translation system. Multilingual ontology, such as Eurovoc, is also popular for MLDC (Pouliquen et al., 2004). However, such resources are scarce and expensive to build. Several other drawbacks of using this technique include dictionary limitation and word ambiguity.

More recently, parallel texts have been used to connect document collections from different languages (Wei et al., 2008). This is done by collapsing columns in a term by document matrix that are translations of each other. Nevertheless, building parallel texts is also expensive and requires a lot of works, hence shifting the paradigm of multilingual works to comparable corpora.

Comparable corpora are collections of texts in different languages regarding similar topics produced at the same time. The key difference between comparable corpora and parallel texts is that documents in comparable corpora are not necessarily translations of each other. They are easier to be acquired, and do not need exhaustive works to be prepared. News agencies often give information in many different languages and can be good sources for comparable corpora. Terms in comparable corpora, being about the same topic, up to some point explain the same concepts in different languages. Pairing comparable corpora with spectral method such as Latent Semantic Analysis has become prevalent, e.g. (Gliozzo and Strapparava, 2005). They rely on the presence of common words and proper nouns among various languages to build a language-independent space. The performance of such method is highly dependent on the languages being used. Here, we present another approach to exploit knowledge in comparable corpora; using propagation method to aid spreading similarity between collections of documents in different languages.

Spectral clustering is the task of finding good clusters by using information contained in the eigenvectors of a matrix derived from the data. It has been successfully applied in many applications including information retrieval (Deerwester et al., 2003) and computer vision (Meila and Shi, 2000). An in-depth analysis of spectral algorithm for clustering problems is given in (Ng et al., 2002). Zhang and Mao (2008) used a related

technique called Modularity Eigenmap to extract community structure features from the document network to solve hypertext classification problem.

Semi-supervised clustering enhances clustering task by incorporating prior knowledge to aid clustering process. It allows user to guide the clustering process by giving some feedback to the model. In traditional clustering algorithm, only unlabeled data is used to find assignments of data points to clusters. In semi-supervised clustering, prior knowledge is given to improve performance of the system. The supervision is usually given as pair of must-linked constraints and cannot link constraints, first introduced in (Wagstaff and Cardie, 2000). Kamvar et al. (2003) proposed spectral learning algorithm that can take supervisory information in the form of pairwise constraints or labeled data. Their algorithm is intended to be used in monolingual context, while our algorithm is designed to work in multilingual context.

## 3 Multilingual Spectral Clustering

There have been several works on multilingual document clustering as mention previously in Section 2. Our key contribution here is the propagation method to make spectral clustering algorithm works for multilingual problems. The clustering model exploits the supervisory information by detecting  $k$  nearest neighbors of the newly-linked documents, and propagates document similarity to these neighbors. The model can be applied to any multilingual text collections regardless of the languages. Overall algorithm is given in Section 3.1 and the method to merge multilingual spaces by similarity propagation is given in Section 3.2.

### 3.1 Spectral Clustering Algorithm

Spectral clustering tries to find good clusters by using top eigenvectors of normalized data affinity matrix. The document set is being modeled as undirected graph  $G(V, E, W)$ , where  $V$ ,  $E$ , and  $W$  denote the graph vertex set, edge set, and transition probability matrix, respectively. In graph  $G$ ,  $v \in V$  represents a document, and weight  $w_{ij} \in W$  represents transition probability between document  $v_i$  to  $v_j$ . The transition probabilities can be interpreted as edge flows in Markov random walk over graph vertices (documents in collections).

Algorithm to perform spectral clustering is given in Algorithm 1. Let  $A$  be affinity matrix

where element  $A_{ij}$  is cosine similarity between document  $v_i$  and  $v_j$  (Algorithm 1, line 1). It is straightforward that documents belonging to different languages will have similarity zero. Rare exception occurs when they have common words because the languages are related one another. As a consequence, the similarity matrix will have many zeros. Our model amplifies prior knowledge in the form of comparable corpora by performing document similarity propagation, presented in Section 3.2 (Algorithm 1, line 4; Algorithm 2, explained in Section 3.2). After propagation, the affinity matrix is post-processed (Algorithm 1, line 6, explained in Section 3.2) before being transformed into transition probability matrix.

The transformation can be done using any normalization for spectral methods. Define  $N = D^{-1}A$ , as in (Meila and Shi, 2001), where  $D$  is the diagonal matrix whose elements  $D_{ij} = \sum_j A_{ij}$  (Algorithm 1, line 7). Alternatively, we can define  $N = D^{-1/2}AD^{-1/2}$  (Ng et al., 2002), or  $N = (A + d_{max}I - D)/d_{max}$  (Fiedler, 1975), where  $d_{max}$  is the maximum rowsum of  $A$ . For our experiment, we use the first normalization method, though other methods can be applied as well.

Meila and Shi (2001) show that probability transition matrix  $N$  with  $t$  strong clusters will have  $t$  piecewise constant eigenvectors. They also suggest using these  $t$  eigenvectors in clustering process. We use the information contains in  $t$  largest eigenvectors of  $N$  (Algorithm 1, line 8-11) and perform  $K$ -means clustering algorithm to find the data clusters (Algorithm 1, line 12).

### 3.2 Propagating Prior Knowledge

We use information obtained from comparable corpora to merge multilingual language spaces. Suppose we have text collections in  $L$  different languages. We combine this collections with comparable corpora, also in  $L$  languages, that act as our supervisory information. Comparable corpora are used to gather prior knowledge by making must-linked constraints for documents in different languages that belong to the same topic in the corpora, propagating similarity to other documents while doing so.

Initially, our affinity matrix  $A$  represents cosine similarity between all pairs of documents.  $A_{ij}$  is set to zero if  $j$  is not the top  $k$  nearest neighbors of  $i$  and likewise. Next, set  $A_{ij}$  and  $A_{ji}$  to 1 if document  $i$  and document  $j$  are different in lan-

---

#### Algorithm 1 Multilingual Spectral Clustering

---

**Input:** Term by document matrix  $M$ , pairwise constraints

**Output:** Document clusters

- 1: Create graph affinity matrix  $A \in \mathbb{R}^{n \times n}$  where each element  $A_{ij}$  represents the similarity between document  $v_i$  and  $v_j$ .
  - 2: **for all** pairwise constraints in comparable corpora **do**
  - 3:    $A_{ij} \leftarrow 1, A_{ji} \leftarrow 1$ .
  - 4:   Recursive Propagation ( $A, S, \beta, k, v_i, v_j$ ).
  - 5: **end for**
  - 6: Post-process matrix  $A$  so that every value in  $A$  is greater than  $\delta$  and less than 1.
  - 7: Form a diagonal matrix  $D$ , where  $D_{ii} = \sum_j A_{ij}$ . Normalize  $N = D^{-1}A$ .
  - 8: Find  $x_1, x_2, \dots, x_t$ , the  $t$  largest eigenvectors of  $N$ .
  - 9: Form matrix  $X = [x_1, x_2, \dots, x_t] \in \mathbb{R}^{n \times t}$ .
  - 10: Normalize row  $X$  to be unit length.
  - 11: Project each document into eigen-space spanned by the above  $t$  eigenvectors (by treating each row of  $X$  as a point in  $\mathbb{R}^t$ , row  $i$  represents document  $v_i$ ).
  - 12: Apply  $K$ -means algorithm in this space to find document clusters.
- 

guage and belong to the same topic in our comparable corpora. This will incorporate the must-linked constraint to our model. We can also give supervisory information for pairs of document in the same language, but this is optional. We also do not use cannot-linked constraints since the main goal is to merge multilingual spaces. In our experiment we show that using only must-linked constraints with propagation is enough to achieve encouraging clustering results.

The supervisory information acquired from comparable corpora only connects two nodes in our graph. Therefore, the number of edges between documents in different languages is about as many as the number of must-linked constraints given. We argue that we need more edges between pairs of documents in different languages to get better results.

We try to build more edges by propagating similarity to other documents that are most similar to the newly-linked documents. Figure 1 gives an illustration of edge-creation process when two multilingual documents (nodes) are connected. Sup-

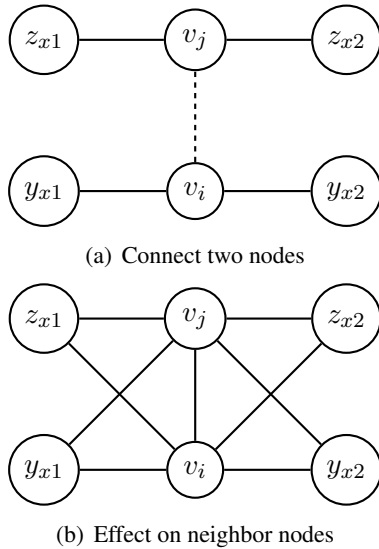


Figure 1: Pairing two multilingual documents affect their neighbors.  $v_i$  and  $v_j$  are documents in two different languages.  $y_x$  and  $z_x$  are neighbors of  $v_i$  and  $v_j$  respectively.

pose that we have six documents in two different languages. Initially, documents are only connected with other documents that belong to the same language. The supervisory information tells us that two multilingual documents  $v_i$  and  $v_j$  should be connected (Figure 1(a)). We then build an edge between these two documents. Furthermore, we also use this information to build edges between  $v_i$  and neighbors of  $v_j$  and likewise (Figure 1(b)).

This follows from the hypothesis that bringing together two documents should also bring other documents that are similar to those two closer in our clustering space. Klein et al. (2002) stated that a good clustering algorithm, besides satisfying known constraints, should also be able to satisfy the implications of those constraints. Here, we allow not only instance-level inductive implications, but utilize it to get higher-level inductive implications. In other words, we alter similarity space so that it can detect other clusters by changing the topology of the original space.

The process is analogous to shortening the distance between sets of documents in Euclidean space. In vector space model, two documents that are close to each other have high similarity, and thus will belong to the same cluster. Pairing two documents can be seen as setting the distance in this space to 0, thus raising their similarity to 1. While doing so, each document would also draw

sets of documents connected to it closer to the centre of the merge, which is equivalent to increasing their similarities.

Suppose we have document  $v_i$  and  $v_j$ , and  $y$  and  $z$  are sets of their respective  $k$  nearest neighbors, where  $|y| = |z| = k$ . The propagation method is a recursive algorithm with base  $S$ , the number of desired level of propagation. Recursive  $k$ -nearest neighbor makes decision to give high similarity between multilingual documents not only determined by their similarity to the newly-linked documents, but also their similarity to the  $k$  nearest neighbors of the respective document. Several documents are affected by a single supervisory information. This will prove useful when only limited amount of supervisory information given. It uses document similarity matrix  $A$ , as defined in the previous section.

1. For  $y_x \in y$  we propagate  $\beta A_{v_i y_x}$  to  $A_{v_j y_x}$ . Set  $A_{y_x v_j} = A_{v_j y_x}$  (Algorithm 2, line 5-6). In other words, we propagate the similarity between document  $v_i$  and  $y$  nearest neighbors of  $v_i$  to document  $v_j$ .
2. Similarly, for  $z_x \in z$  we propagate  $\beta A_{v_j z_x}$  to  $A_{v_i z_x}$ . Set  $A_{z_x v_i} = A_{v_i z_x}$  (Algorithm 2, line 10-11). In other words, we propagate the similarity between document  $v_j$  and  $z$  nearest neighbors of  $v_j$  to document  $v_i$ .
3. Propagate higher order similarity to  $k$  nearest neighbors of  $y$  and  $z$ , discounting the similarity quadratically, until required level of propagation  $S$  is reached (Algorithm 2, line 7 and 12).

The coefficient  $\beta$  represents the degree of enforcement that the documents similar to a document in one language, will also have high similarity with other document in other language that is paired up with its ancestor. On the other hand,  $k$  represents the number of documents that are affected by pairing up two multilingual documents. After propagation, similarity of documents that falls below some threshold  $\delta$  is set to zero (Algorithm 1, line 6). This post-processing step is performed to nullify insignificant similarity values propagated to a document. Additionally, if there exists similarity of documents that is higher than one, it is set to one.

---

**Algorithm 2** Recursive Propagation

---

**Input:** Affinity matrix  $A$ , level of propagation  $S$ ,  $\beta$ , number of nearest neighbors  $k$ , document  $v_i$  and  $v_j$

**Output:** Propagated affinity matrix

```
1: if  $S = 0$  then
2:   return
3: else
4:   for all  $y_x \in k$ -NN document  $v_i$  do
5:      $A_{v_j y_x} \leftarrow A_{v_j y_x} + \beta A_{v_i y_x}$ 
6:      $A_{y_x v_j} \leftarrow A_{v_j y_x}$ 
7:     Recursive Propagation ( $A, S - 1,$ 
8:        $\beta^2, k, y_x, v_j$ )
9:   end for
10:  for all  $z_x \in k$ -NN document  $v_j$  do
11:    Set  $A_{v_i z_x} \leftarrow A_{v_i z_x} + \beta A_{v_j z_x}$ 
12:    Set  $A_{z_x v_i} \leftarrow A_{v_i z_x}$ 
13:    Recursive Propagation ( $A, S - 1,$ 
14:       $\beta^2, k, v_i, z_x$ )
15:  end for
16: end if
```

---

## 4 Performance Evaluation

The goals of empirical evaluation include (1) testing whether the propagation method can merge multilingual space and produce acceptable clustering results; (2) comparing the performance to spectral clustering method without propagation.

### 4.1 Data Description

We tested our model using Reuters Corpus Volume 2 (RCV2), a multilingual corpus containing news in thirteen different languages. For our experiment, three different languages: English, French, and Spanish; in six different topics: science, sports, disasters accidents, religion, health, and economy are used. We discarded documents with multiple category labels.

We do not apply any language specific preprocessing method to the raw text data. Monolingual TFIDF is used for feature weighting. All document vectors are then converted into unit vector by dividing by its length. Table 1 shows the average length of documents in our corpus.

### 4.2 Evaluation Metric

For our experiment, we used Rand Index (RI) which is a common evaluation technique for clustering task where the true class of unlabeled data

	English	French	Spanish	Total
<b>Science</b>	290.10	165.10	213.45	222.88
<b>Sports</b>	182.55	156.83	189.75	176.37
<b>Disasters</b>	154.29	175.89	165.31	165.16
<b>Religion</b>	317.77	177.91	242.67	246.11
<b>Health</b>	251.19	233.70	227.25	237.38
<b>Economy</b>	266.89	192.55	306.11	255.08
<b>Total</b>	243.79	183.61	224.09	217.16

Table 1: Average number of words of documents in the corpus. Each language consists of 600 documents, and each topic consists of 100 documents (per language).

is known. Rand Index measures the percentage of decisions that are correct, or simply the accuracy of the model. Rand Index is defined as:

$$RI = \frac{TP + TN}{TP + FP + TN + FN}$$

Rand Index penalizes false positive and false negative decisions during clustering. It takes into account decision that assign two similar documents to one cluster (TP), two dissimilar documents to different clusters (TN), two similar documents to different clusters (FN), and two dissimilar documents to one cluster (FP). We do not include links created by supervisory information when calculating true positive decisions and only consider the number of free decisions made.

We also used  $F_\alpha$ -measure, the weighted harmonic mean of precision (P) and recall (R).  $F_\alpha$ -measure is defined as:

$$F_\alpha = \frac{(\alpha^2 + 1)PR}{\alpha^2 P + R}$$
$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$

Last, we used purity to evaluate the accuracy of assignments. Purity is defined as:

$$Purity = \frac{1}{N} \sum_t \max_j |\omega_t \cap c_j|$$

where  $N$  is the number of documents,  $t$  is the number of clusters,  $j$  is the number of classes,  $\omega_t$  and  $c_j$  are sets of documents in cluster  $t$  and class  $j$  respectively.

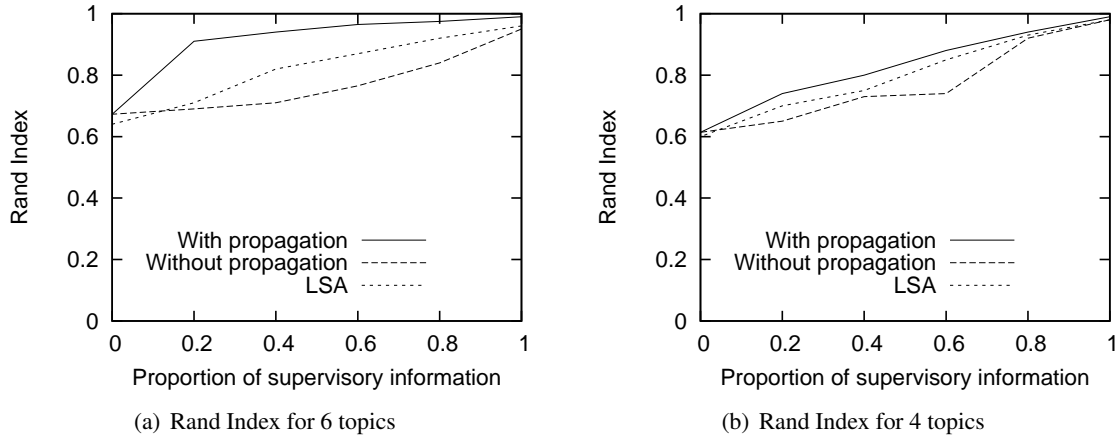


Figure 2: Rand Index on the RCV2 task with (a) 1800 documents, 6 topics; and (b) 1200 documents, 4 topics as the proportion of supervisory information increases.  $k = 30$ ,  $\delta = 0.03$ ,  $\beta = 0.5$ ,  $t =$  number of topics, and  $S = 2$ .

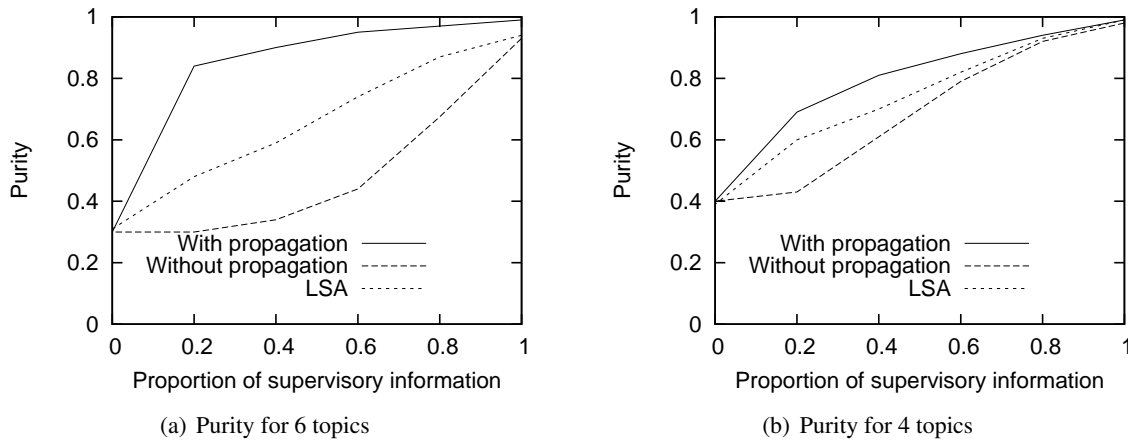


Figure 3: Purity on the RCV2 task with (a) 1800 documents, 6 topics; and (b) 1200 documents, 4 topics as the proportion of supervisory information increases.  $k = 30$ ,  $\delta = 0.03$ ,  $\beta = 0.5$ ,  $t =$  number of topics, and  $S = 2$ .

### 4.3 Experimental Results

To prove the effectiveness of our clustering algorithm, we performed the following experiments on our data set. We first tested our algorithm on four topics, science, sports, religion, and economy. We then tested our algorithm using all six topics to get an understanding of the performance of our model in larger collections with more topics. We used subset of our data as supervisory information and built must-linked constraints from it. The proportion of supervisory information provided to the system is given in  $x$ -axis (Figure 2 - Figure 4.3). 0.2 here means 20% of documents in each language are taken to be used as prior knowledge. Since the number of documents in each language for our experiment is the same, we have the same

numbers of documents in subset of English collection, subset of French collection, and subset of Spanish collection. We also ensure there are same numbers of documents for a particular topic in all three languages. We can build must-linked constraints as follows. For each document in the subset of English collection, we create must-linked constraints with one randomly selected document from the subset of French collection and one randomly selected document from the subset of Spanish collection that belong to the same topic with it. We then create must-linked constraint between the respective French and Spanish documents. The constraints given to the algorithm are chosen so that there are several links that connect every topic in every language. Note that the class label in-

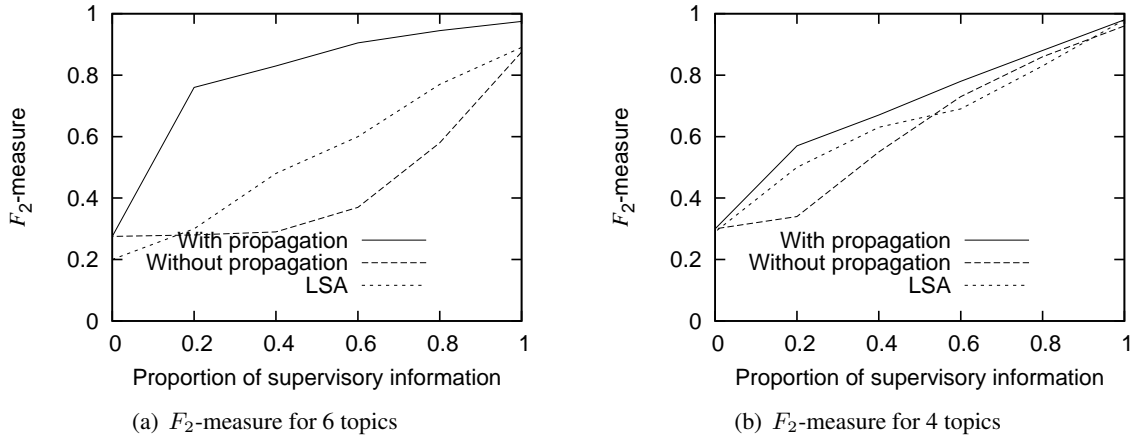


Figure 4:  $F_2$ -measure on the RCV2 task with (a) 1800 documents, 6 topics; and (b) 1200 documents, 4 topics as the proportion of supervisory information increases.  $k = 30$ ,  $\delta = 0.03$ ,  $\beta = 0.5$ ,  $t = \text{number of topics}$ , and  $S = 2$ .

formation is only used to build must-linked constraints between documents, and we do not assign the documents to a particular cluster.

Figure 2 shows the Rand Index as proportion of supervisory information increases. Figure 3 and Figure 4.3 give purity and  $F_2$ -measure for the algorithm respectively. To show the importance of the propagation in multilingual space, we give comparison with spectral clustering model without propagation. Three lines in Figure 2 to Figure 4.3 indicate: (1) results with propagation (solid line); (2) results without propagation (long-dashed line); and (3) results using Latent Semantic Analysis(LSA)-based method by exploiting common words between languages (short-dashed line). For each figure, 6 plots are taken starting from 0 in 0.2-point-increments. We conducted the experiments three times for each proportion of supervisory information and use the average values. As we can see from Figure 2, Figure 3, and Figure 4.3, the propagation method can significantly improve the performance of spectral clustering algorithm. For 1800 documents in 6 topics, we manage to achieve  $RI = 0.91$ , purity = 0.84, and  $F_2$ -measure = 0.76 with only 20% of documents (360 documents) used as supervisory information. Spectral clustering algorithm without propagation can only achieve 0.69, 0.30, 0.28 for RI, purity, and  $F_2$ -measure respectively. The propagation method is highly effective when only small amount of supervisory information given to the algorithm. Obviously, the more supervisory information given, the better the performance is. As the number of supervisory information increases,

the difference of the model performance with and without propagation becomes smaller. This is because there are already enough links between multilingual documents, so we do not necessarily build more links through similarity propagation anymore. However, even when there are already many links, our model with propagation still outperforms the model without propagation.

We compare the performance of our algorithm to LSA-based multilingual document clustering model. We performed LSA to the multilingual term by document matrix. We do not use parallel texts and only rely on common words across languages as well as must-linked constraints to build multilingual space. The results show that exploiting common words between languages alone is not enough to build a good multilingual semantic space, justifying the usage of supervisory information in multilingual document clustering task. When supervisory information is introduced, our method achieves better results than LSA-based method. In general, the LSA-based method performs better than the model without propagation.

We assess the sensitivity of our algorithm to parameter  $\beta$ , the penalty for similarity propagation. We assess the sensitivity of our algorithm to parameter  $\beta$ , the penalty for similarity propagation. We tested our algorithm using various  $\beta$ , starting from 0 to 1 in 0.2-point-increments, while other parameters being held constant. Figure 5(a) shows that changing  $\beta$  to some extent affects the performance of the algorithm. However, after some value of reasonable  $\beta$  is found, increasing  $\beta$  does not have significant impact on the per-

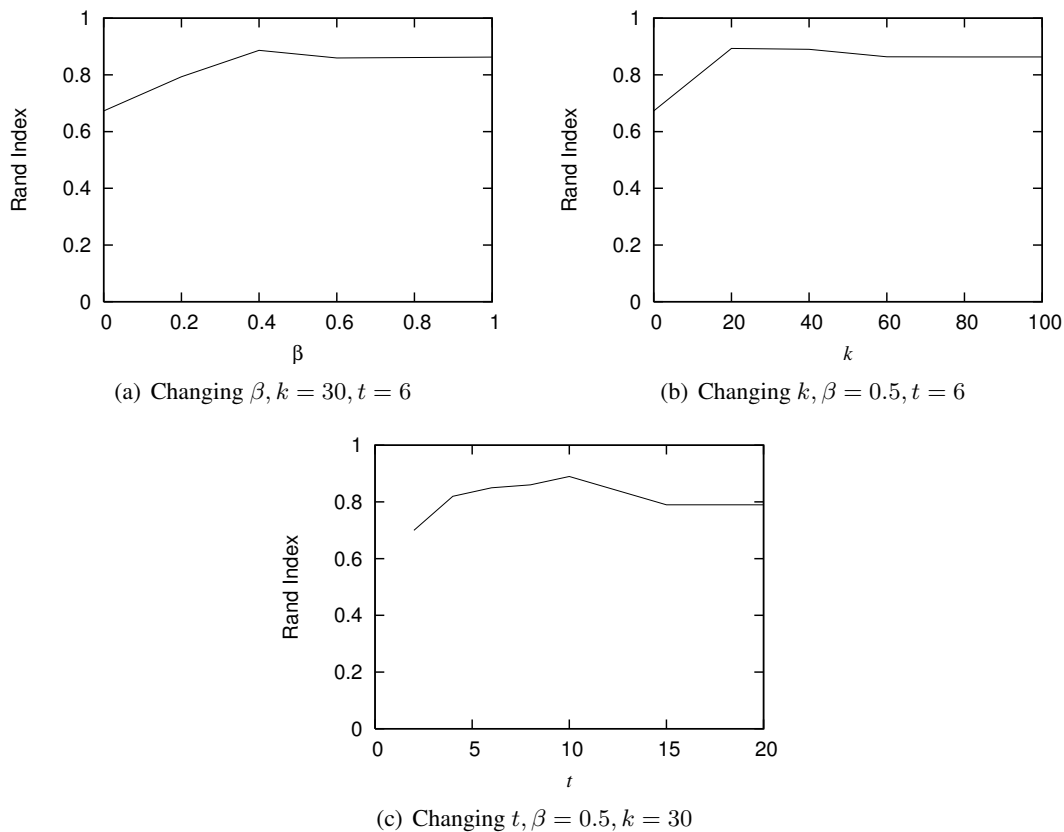


Figure 5: Rand Index on the RCV2 task with 1800 documents and 6 topics as (a)  $\beta$  increases; (b)  $k$  increases; and (c)  $t$  increases.  $\delta = 0.03$ ,  $S = 2$ , and 20% of documents are used as supervisory information.

formance of the algorithm. We also tested our algorithm using various  $k$ , starting from 0 to 100 in 20-point-increments. Figure 5(b) reveals that the performances of the model with different  $k$  are comparable, as long as  $k$  is not too small. However, using too large  $k$  will slightly decrease the performance of the model. Too many propagations make several dissimilar documents receive high similarity value that cannot be nullified by the post-processing step. Last, we experimented using various  $t$  ranging from 2 to 20. Figure 5(c) shows that the method performs best when  $t = 10$ , and for reasonable value of  $t$  the method achieves comparable performance.

## 5 Conclusion

We present here a multilingual spectral clustering model that is able to work irrespective of the languages being used. The key component of our model is the propagation algorithm to merge multilingual spaces. We tested our algorithm on Reuters RCV2 Corpus and compared the performance with spectral clustering model without

propagation. Experimental results reveal that using limited supervisory information, the algorithm achieves encouraging clustering results.

## References

- Charu C. Aggarwal, Stephen C. Gates and Philip S. Yu. 1999. On The Merits of Building Categorization Systems by Supervised Clustering. In *Proceedings of Conference on Knowledge Discovery in Databases*:352-356.
- Hsin-Hsi Chen and Chuan-Jie Lin. 2000. A Multilingual News Summarizer. In *Proceedings of 18th International Conference on Computational Linguistics*:159-165.
- Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*:41(6):391-407.
- Miroslav Fiedler. 1975. A Property of Eigenvectors of Nonnegative Symmetric Matrices and its Applications to Graph Theory. *Czechoslovak Mathematical Journal*, 25:619-672.



- Alfio Gliozzo and Carlo Strapparava. 2005. Cross language Text Categorization by acquiring Multilingual Domain Models from Comparable Corpora. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*:9-16.
- Sepandar D. Kamvar, Dan Klein, and Christopher D. Manning. 2003. Spectral Learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.
- Dan Klein, Sepandar D. Kamvar, and Christopher D. Manning. 2002. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *The Nineteenth International Conference on Machine Learning*.
- Xiaoyong Liu and W. Bruce Croft. 2004. Cluster-based Retrieval using Language Models. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*:186-193.
- Marinla Meilă and Jianbo Shi. 2000. Learning segmentation by random walks. In *Advances in Neural Information Processing Systems*:873-879.
- Marinla Meilă and Jianbo Shi. 2001. A Random Walks View of Spectral Segmentation. In *AI and Statistics (AISTATS)*.
- Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. 2002. On Spectral Clustering: Analysis and an algorithm. In *Proceedings of Advances in Neural Information Processing Systems (NIPS 14)*.
- Bruno Pouliquen, Ralf Steinberger, Camelia Ignat, Emilia Käsper, and Irina Temnikova. 2004. Multilingual and Cross-lingual News Topic Tracking. In *Proceedings of the 20th International Conference on Computational Linguistics*.
- Stefan Siersdorfer and Sergej Sizov. 2004. Restrictive Clustering and Metaclustering for Self-Organizing Document. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*.
- Kiri Wagstaff and Claire Cardie 2000. Clustering with Instance-level Constraints. In *Proceedings of the 17th International Conference on Machine Learning*:1103-1110.
- Chih-Ping Wei, Christopher C. Yang, and Chia-Min Lin. 2008. A Latent Semantic Indexing Based Approach to Multilingual Document Clustering. In *Decision Support Systems*, 45(3):606-620
- Dell Zhang and Robert Mao. 2008. Extracting Community Structure Features for Hypertext Classification. In *Proceedings of the 3rd IEEE International Conference on Digital Information Management (ICDIM)*.