

Automatic Relation Extraction with Model Order Selection and Discriminative Label Identification

Chen Jinxiu¹, Ji Donghong¹, Tan Chew Lim², and Niu Zhengyu¹

¹ Institute of Infocomm Research,
21 Heng Mui Keng Terrace, Singapore, 119613
{jinxiu, dhji, zniu}@i2r.a-star.edu.sg

² Department of Computer Science,
National University of Singapore, Singapore, 117543
tancl@comp.nus.edu.sg

Abstract. In this paper, we study the problem of unsupervised relation extraction based on model order identification and discriminative feature analysis. The model order identification is achieved by stability-based clustering and used to infer the number of the relation types between entity pairs automatically. The discriminative feature analysis is used to find discriminative feature words to name the relation types. Experiments on ACE corpus show that the method is promising.

1 Introduction

Relation extraction is the task of finding relationships between two entities from text contents. Recently, it has received more and more attention in many areas, e.g., information extraction, ontology construction, and bioinformatics, etc. In this paper, we propose an unsupervised method for relation extraction from corpus.

Since the concept of relation extraction was introduced in MUC 6 [1], there has been considerable work on supervised learning of relation patterns, using corpora which have been annotated to indicate the information to be extracted [2,9,8]. A range of extraction models have been used, including both symbolic rules and statistical rules such as HMMs or Kernels. These methods have been particularly successful in some specific domains. However, manually tagging of large amounts of training data is very time-consuming; furthermore, it is difficult for one extraction system to be ported across different domains.

Due to the limitation of supervised methods, some weakly (semi-) supervised approaches have been suggested [3,6,5,4]. One common feature of these algorithms is that they need to pre-define some initial seeds for any particular relation, then bootstrap from the seeds to acquire the relation. However, to determine how to select these seeds and how many seeds to be selected tends to be very subjective.

Hasegawa, et al. put forward an unsupervised approach for relation extraction from large text corpora [7]. Their assumption is that pairs of entities with

same relation between them tend to occur in similar contexts, and the representative words in the contexts can be regarded as somewhat characterization of the relation. Thus their method contains two key steps, the first is to cluster the contexts in which the pairs of entities occur, and the second is to extract the representative words from the contexts.

For the unsupervised approach, we noticed some limitations. First, they adopted a hierarchical clustering method to cluster the contexts. However, the similarity threshold for the clusters, like the appropriate number of clusters, is somewhat difficult to pre-define. Second, after context clustering, they selected the most frequent words in the contexts to represent the relation that holds between the entities. However, such words may occur frequently in any other clusters too. Hence, they may not have quality to discriminate between clusters.

In this paper, we try to resolve the above limitations of the unsupervised approach by model order selection and discriminative label identification. First, we adopt a stability-based method to cluster the contexts, which can infer the number of the appropriate clusters automatically. Second, we propose a feature weighting method and try to extract more discriminative words from the contexts to represent the relations.

The rest of this paper is organized as follows. In section 2 we overview the main phases in our proposed method. In section 3 we present the stability based main analysis algorithm to estimate the “correct” number of relation types. In section 4 we talk about how to identify discriminative labels for each relation type. Then we describe experiments and evaluations in section 5. In section 6 we give some discussions about our approach. Finally, the conclusions and future work are given in section 7.

2 Overview

For each pair of entities (E_1 and E_2)¹ with at least one known relation, we propose a method to cluster these relations into similar types. To discover such relationships, our proposed approach consists of the following three phases.

- to collect the contexts in which the entities co-occur;
- to cluster the contexts using stability-based method;
- to select discriminative features and label the clusters.

In phase 1, we assume that for any two particulate entities $e_1 \in E_1$, and $e_2 \in E_2$, they may hold more than one kind of relations. So, we collect the contexts from a corpus in which e_1 and e_2 co-occur within a context window of d words. Here, the context includes the words between, before and after them (In this paper, we use only words as the features of context vectors.). In fact, the approach also applies to the cases that e_1 and e_2 hold only one kind of relations, in such cases, we need to collect and accumulate the contexts.

¹ Entities refer to names, location, proper nouns, etc.

In phase 2, we cluster each context by the type of relation it represents. For a cluster c with a relation r , the entities e_1 and e_2 whose context belongs to c can be regarded as holding the relation r . In our experiments, cosine-similarity measure is used to compare contexts.

In phase 3, we select feature words from the contexts for each cluster as the label of the cluster, and it can also be seen as the name of the relation type. So the feature word should be discriminative among the clusters.

3 Context Clustering

Since we do not know how many relation types in advance and do not have any labelled relation training examples at hand, the problem of model order selection arises, i.e. estimating the “correct” number of clusters. In this paper, the model selection capability is achieved by resampling based stability analysis, which has been successfully applied to several unsupervised learning problems (e.g. [11], [10], [13], [12]).

To estimate the number of the clusters, we need a criterion to evaluate the merit for each possible number of clusters, and select the model order which maximizes the criterion. Formally, let k be the model order, we need to find k in Equation: $k = \arg \max_k \{criterion(k)\}$. Here, the criterion is set up based on resampling-based stability.

Let P^μ be a subset sampled from full entity pairs set P with size $\alpha|P|$ (α set as 0.9 in this paper.), $C(C^\mu)$ be $|P| \times |P|$ ($|P^\mu| \times |P^\mu|$) connectivity matrix based on the clustering results on $P(P^\mu)$. Each entry $c_{ij}(c_{ij}^\mu)$ of $C(C^\mu)$ is calculated in the following: if the entity pair $p_i \in P(P^\mu)$, $p_j \in P(P^\mu)$ belong to the same cluster, then $c_{ij}(c_{ij}^\mu)$ equals 1, else 0. Then the stability is defined in Equation 1:

$$M(C^\mu, C) = \frac{\sum_{i,j} 1\{C_{i,j}^\mu = C_{i,j} = 1, p_i \in P^\mu, p_j \in P^\mu\}}{\sum_{i,j} 1\{C_{i,j} = 1, p_i \in P^\mu, p_j \in P^\mu\}} \quad (1)$$

Intuitively, $M(C^\mu, C)$ denotes the consistency between the clustering results on C^μ and C . The assumption is that if the cluster number k is actually the “natural” number of relation types, then clustering results on subsets P^μ generated by sampling should be similar to the clustering result on full entity pair set P . Obviously, the above function satisfies $0 \leq M \leq 1$.

It is noticed that $M(C^\mu, C)$ tends to decrease when increasing the value of k . Therefore for avoiding the bias that small value of k is to be selected as cluster number, we use the cluster validity of a random predictor ρ_k to normalize $M(C^\mu, C)$. The random predictor ρ_k achieved the stability value by assigning uniformly drawn labels to objects, that is, splitting the data into k clusters randomly. Furthermore, for each k , we tried q times. So, the normalized object function can be defined as equations 2:

$$M_k^{norm} = \frac{1}{q} \sum_{i=1}^q M(C_k^{\mu_i}, C_k) - \frac{1}{q} \sum_{i=1}^q M(C_{\rho_k}^{\mu_i}, C_{\rho_k}) \quad (2)$$

Normalizing $M(C^\mu, C)$ by the stability of the random predictor can yield values independent of k . The effect of such normalization can be observed from the experimental results (See Table 5). The overall algorithm is in Table 1. Table 2 shows the evaluation procedure of model order selection.

Table 1. Model Selection Algorithm for Relation Extraction

Input: Corpus D tagged with Entities(E_1, E_2);
Output: Model Order (number of relation types);

1. Collect the contexts of all entity pairs in the document corpus D , namely P ;
2. Set the range (K_l, K_h) for the possible number of relation clusters;
3. Set estimated model order $k = K_l$;
4. Cluster all entity pairs set P into k clusters using stability analysis method;
5. Record k and the score of the merit of k , namely M_k ;
6. If $k < K_h$, $k = k + 1$, go to step 4; otherwise, go to Step 7;
7. Select k which maximizes the score of the merit M_k ;

Table 2. Unsupervised Algorithm for Evaluation of Model Order Selection

Function: criterion(k, P, q)
Input: cluster number k , entity pairs set P , and sampling frequency q ;
Output: the score of the merit of k ;

1. With k as input, perform k -means clustering analysis on pairs set P ;
2. Construct connectivity matrix C_k based on above clustering solution on P ;
3. Use random predictor ρ_k to assign uniformly drawn labels to each object in P ;
4. Construct connectivity matrix C_{ρ_k} based on above clustering solution on P ;
5. Construct q subsets of the full pairs set, by randomly selecting αN of the N original pairs, $0 \leq \alpha \leq 1$;
6. For each subset, perform the clustering analysis in Step 2, 3, 4, and result $C_k^\mu, C_{\rho_k}^\mu$;
7. Compute M_k to evaluate the merit of k using Equation 2;
8. Return M_k ;

Table 3. Some Context examples in two clusters of the output in the domain PER-ORG

Cluster 1:

[PER] vice president of the [ORG]
[PER] president and chief operating officer of [ORG]
[PER] senior vice president of [ORG]
⋮

Cluster 2:

[PER] joined the communist -backed [ORG]
[PER] and joined a laborer's [ORG]
[PER] a partner in Blackstone, will join Host Marriott's [ORG]
⋮

After the number of optimal clusters has been chosen, we adopted the kmeans algorithm for the clustering phase. The output of context clustering is a set of context clusters, each of them is supposed to denote one relation type. As an example, Table 3 lists two clusters with some context examples.

4 Relation Labelling

For labelling each relation type, we use DCM (discriminative category matching) scheme to identify discriminative label, which is also used in document classification [14] and weights the importance of a feature based on their distribution. In this scheme, a feature is not important if the feature appears in many clusters and is evenly distributed in these clusters, otherwise it will be assigned higher importance.

To weight a feature f_i within a category, we take into account the following information to ensure the selected features have the discrimination power:

- The relative importance of f_i within a cluster:

$$WC_{i,k} = \frac{\log_2(pf_{i,k} + 1)}{\log_2(N_k + 1)}. \quad (3)$$

where $pf_{i,k}$ is the number of those term pairs which contain feature f_i in cluster k . N_k is the total number of term pairs in cluster k .

- The relative importance of f_i across clusters:

$$CC_i = \log \frac{N \cdot \max_{k \in C_i} \{WC_{i,k}\}}{\sum_{k=1}^N WC_{i,k}} \cdot \frac{1}{\log N} \quad (4)$$

where C_i is the set of clusters which contain feature f_i . N is the total number of clusters.

In Equation 3, both numerator and denominator are logarithmic, which is based on the observation that the frequency of a feature appearing over many pairs is rare.

In Equation 4, the summation is used for gathering the total importance of a feature across all categories, while the maximum is used for averaging the summation value. If a feature is regarded as important in many clusters, then this feature is obviously not important for relation labelling. In other words, the higher the value of the numerator, the smaller the discriminative power is. The term, $1/\log(N)$, is used for normalization such that $0 \leq CC_i \leq 1$.

Here, $WC_{i,k}$ and CC_i are designed to capture the local information within a cluster and global information about the feature distribution across clusters respectively. It is insufficient to describe the importance and discrimination power of a feature by examining any one information only. As a result, combining both $WC_{i,k}$ and CC_i we define the weight $W_{i,k}$ of f_i in cluster k as:

$$W_{i,k} = \frac{WC_{i,k}^2 \cdot CC_i^2}{\sqrt{WC_{i,k}^2 + CC_i^2}} \cdot \sqrt{2} \quad (5)$$

Where $\sqrt{2}$ is used for normalization such that $0 \leq W_{i,k} \leq 1$.

As an example, we list some features with their weighting score for the two clusters in Table 3.

Cluster 1: (president, 0.629374), (chairman, 0.188562), (ceo, 0.155908), (chief, 0.147443), ...;

Cluster 2: (join, 0.272382), (serve, 0.114423), (communist, 0.049031), (bond, 0.044387), (aid, 0.044387), ...

5 Experimental Evaluations

5.1 Data

We constructed three subsets from ACE corpus² for domains PER-ORG (person-organization), ORG-GPE (organization-gpe) and ORG-ORG (organization - organization) respectively. The details of these subsets are given in Table 4, which are broken down by different relation types.

Table 4. Three domains of entity pairs: frequency distribution for different relation types

PER-ORG num:786		ORG-GPE num:262		ORG-ORG num:580	
Relation types	Percentage	Relation types	Percentage	Relation types	Percentage
Management	36.39%	Based-In	46.56%	Member	27.76%
General-staff	29.90%	Located	35.11%	Subsidiary	19.83%
Member	19.34%	Member	11.07%	Part-Of	18.79%
Owner	4.45%	Affiliate-Partner	3.44%	Affiliate-Partner	17.93%
Located	3.28%	Part-Of	2.29%	Owner	8.79%
Client	1.91%	Owner	1.53%	Client	2.59%
Other	1.91%			Management	2.59%
Affiliate-Partner	1.53%			Other	1.21%
Founder	0.76%			Other	0.52%

The ACE corpus contains about 519 files from sources including broadcast, newswire, and newspaper. To verify our proposed method, we only extracted those pairs of entity mentions which have been tagged relation types in the given corpus. Then the relation type tags were removed to test the unsupervised relation disambiguation. During the evaluation procedure, the relation type tags were used as ground truth classes.

The data preprocessing involves lowering the upper case characters, ignoring all words that contain digits or non alpha-numeric characters, removing words from a stop word list, stemming and filtering out low frequency words which appeared only once in the entire set.

² <http://www ldc.upenn.edu/Projects/ACE/>, RDC(Relation Detection and Characterization) is one of the tasks of ACE program. There are five entity types in the ACE corpus, that is, ORGANIZATION, PERSON, GPE, LOCATION and FACILITY.

5.2 Evaluation Method for Clustering Result

When assessing the agreement between clustering result and hand-tagged relation types (ground truth classes), we would encounter the problem that there was no relation type tags for each cluster in our clustering results.

To resolve the problem, we adopted a permutation procedure to assign different relation type tags to only $\min(|EC|, |TC|)$ clusters, where $|EC|$ is the estimated number of clusters, and $|TC|$ is the number of ground truth classes (relation types). This procedure aims to find an one-to-one mapping function Ω from the TC to EC which is based on the assumption that for any two clusters, they do not share the same class labels. Under this assumption, there are at most $|TC|$ clusters which are assigned relation type tags. If the number of the estimated clusters is less than the number of the ground truth clusters, empty clusters should be added so that $|EC| = |TC|$ and the one-to-one mapping can be performed.

With the estimated clusters and the ground truth classes, we construct a contingency table T , where each entry $t_{i,j}$ gives the number of the instances that belong to both the i -th cluster and j -th ground truth class. The mapping procedure can be formulated as the function: $\hat{\Omega} = \arg \max_{\Omega} \sum_{j=1}^{|TC|} t_{\Omega(j),j}$, where $\Omega(j)$ is the index of the estimated cluster associated with the j -th class.

Given the result of one-to-one mapping, we can define the evaluation measure as follows:

$$Accuracy(P) = \frac{\sum_j t_{\hat{\Omega}(j),j}}{\sum_{i,j} t_{i,j}}. \quad (6)$$

Intuitively, it reflects the accuracy of the clustering result.

5.3 Evaluation Method for Relation Labelling

For evaluation of the relation labeling, we need to explore the relatedness between the identified labels and the pre-defined relation names. To do this, we use one information-content based measure [16,15], which is provided in Wordnet-Similarity package [17] to evaluate the similarity between two concepts in Wordnet. Intuitively, the relatedness between two concepts in Wordnet is captured by the information content of their lowest common subsumer (*lcs*) and the information content of the two concepts themselves. This can be viewed as taking the information content of the intersection, which can be formalized as follows:

$$Relatedness_{lin}(c_1, c_2) = \frac{2 \times IC(lcs(c_1, c_2))}{IC(c_1) + IC(c_2)} \quad (7)$$

This measure depends upon the corpus to estimate information content. Information content of a concept is estimated by counting the frequency of that concept in a large corpus and thereby determining its probability via a maximum likelihood estimate. We carried out the experiments using the British National Corpus (BNC) as the source of information content.

5.4 Experiments and Results

For comparison of the effect of the outer and within context of entity pairs, we conducted five different settings of context window size (WIN_{pre} - WIN_{mid} - WIN_{post}) for each domain. For example, the setting of “2-5-2” means that the intervening words between an entity pair should not exceed 5 words and these intervening words together with the two words before the first entity and two words following the second entity constitute the context of an entity pair. Table 5 shows the results of model order identification with unnormalized and normalized objective functions. The results show that the model order identification algorithm with M_k^{unnorm} fail to identify the real number of relation types since the score of M_k^{unnorm} decreased when increasing the cluster number k and finally resulted in 2 clusters over all domains. From Table 5, we can find that with the context setting, 0-10-0, the estimated number of the clusters equals or very closes to the number of classes. It demonstrates that the intervening words less than 10 are appropriate features to reflect the structure behind the contexts, while the intervening words less than 5 is not enough to infer the structure. For the contextual words beyond (before or after) the entities, they tend to be noisy features for the relation estimation, as can be seen that the performance deteriorates when taking them into consideration.

Table 6 shows the accuracy result of the clustering algorithm over three domains with different context window size settings. In this table, we compared

Table 5. Automatically determined the number of relation types using different evaluation functions

Context Window	PER-ORG			ORG-GPE			ORG-ORG		
	Real #	M_k^{unnorm}	M_k^{norm}	Real #	M_k^{unnorm}	M_k^{norm}	Real #	M_k^{unnorm}	M_k^{norm}
0-5-0	9	2	7	6	2	3	9	2	7
2-5-2	9	2	8	6	2	2	9	2	7
0-10-0	9	2	8	6	2	6	9	2	9
2-10-2	9	2	6	6	2	4	9	2	6
5-10-5	9	2	5	6	2	2	9	2	8

Table 6. Performance of the clustering algorithm with various context window size settings over three domains

Context Window size	PER-ORG		ORG-GPE		ORG-ORG	
	Accuracy1 (our method)	Accuracy2 (Hasegawa's method)	Accuracy1 (our method)	Accuracy2 (Hasegawa's method)	Accuracy1 (our method)	Accuracy2 (Hasegawa's method)
0-5-0	33.8%	30.6%	47.4%	47.9%	40.7%	27.2%
2-5-2	35.7%	33.5%	45.2%	43.1%	37.3%	27.6%
0-10-0	39.3%	36.6%	50.9%	42.3%	37.2%	26.3%
2-10-2	32.5%	33.5%	47.8%	43.1%	33.5%	25.3%
5-10-5	30.2%	27.6%	45.7%	42.3%	32.4%	26.0%

our clustering results with the Hasegawa’s clustering algorithm provided in the paper [7], where we specify the cluster number as the number of ground truth classes. Comparing the accuracy of two clustering methods, we can find that our proposed method can achieve better or comparable performance. In addition, in three domains, from the column of accuracy1 with different setting of context window size: (0 – 5 – 0, 2 – 5 – 2), (0 – 10 – 0, 2 – 10 – 2, and 5 – 10 – 5), we can see that the performance does not improve or even becomes worse when extending the context window. The reason is that extending the context may include more features, but at the same time, the noise also increases.

The automatically estimated labels for relation types over 3 domains from estimated clusters are given in Table 7. In each domain, we select two features as labels of each relation type according to their DCM weight scores and calculate the average relatedness between our selected labels (E) and the predefined labels (H). Following the same strategy, we also extracted relation labels (T) from the ground truth classes and provided the average relatedness between T and H. From the column of relatedness (E-H), we can see that it is not easy to find the hand-tagged relation labels exactly, furthermore, the identified labels from the

Table 7. Result for Relation Labelling using DCM strategy. Here, (H) denotes the hand-tagged relation label. (T) denotes the identified relation labels from ground truth classes. (E) is the identified relation labels from our estimated clusters.

Domain	Hand-tagged Label (H)	Identified Label(T)	Identified Label(E)	Relatedness: Ave (T-H)	Relatedness: Ave (E-H)	Relatedness: Ave (E-T)
PER-ORG	management	head,president	president,chairman	0.3703	0.1445	0.8639
	general-staff	work,fire	work,charge	0.6254	0.6411	0.6900
	member	join,communist	join,serve	0.394	0.1681	0.5306
	owner	bond,bought	control,house	0.1351	0.1578	0.4308
	located	appear,include	lobby,appear	0.0000	0.1606	0.2500
	client	hire,reader	bought,consult	0.4378	0.0000	0.1417
	affiliate	affiliate,associate	affiliate,director	0.9118	0.8002	0.8615
	founder	form,found	state,party	0.1516	0.2846	0.3909
ORG-GPE	based-in	base,unit	unit,base	0.7938	0.7938	0.7938
	located	northwest,travel	travel,bank	0.0000	0.1832	0.4734
	member	federal,led	construct,federal	0.2155	0.1408	0.2500
	affiliate	fund,sell	fund,have	0.4779	0.4505	0.6886
	part-of	include,involve	hold,assembly	0.1905	0.6049	0.8893
	owner	represent,mountain	represent,detain	0.0694	0.0000	0.3314
ORG-ORG	member	communist,join	communist,family	0.3141	0.4589	0.6243
	subsidiary	hire,join	agree,form	0.1446	0.1746	0.2406
	part-of	part,include	hold,include	0.6905	0.6137	0.7959
	affiliate	work,affiliate	affiliate,join	0.6719	0.7536	0.6719
	owner	have,asset	asset,deal	0.1782	0.0698	0.7204
	client	service,buy	import,consume	0.1633	0.1377	0.2926
	management	share,control	lead,control	0.3364	0.3146	0.7189

Table 8. Result for Relation Labelling using the frequency strategy in [7] for the domain PER-ORG

Hand-tagged Relation Label (H)	Identified Relation Label from ground truth classes (T)	Identified Relation Label from estimated clusters (E)
management	said,chief	two,head
general-staff	said,work	work,said
member	said,two	say,join
owner	bond,bought	are,become
located	said,appear	are,appear
client	hire,reader	said,bought
affiliate-partner	affiliate,associate	affiliate,three
founder	form,found	work,said

ground-truth classes are either non-comparable to the pre-defined labels in most cases (T-H). The reason may be that the pre-defined relation names tend to be some abstract labels over the features, e.g., ‘management’ vs. ‘president’, ‘head’ or ‘chairman’; ‘member’ vs. ‘join’, ‘serve’, etc., while the abstract words and the features are located far away in Wordnet.

Table 7 also lists the relatedness between the labels identified from the clusters and those from ground truth classes (E-T). We can see that the labels are comparable by their average relatedness.

Table 8 shows the result for relation labelling using frequency strategy in [7] for the domain PER-ORG. From this table, we can see that the frequency strategy is likely to select those common words as the relation labels, which may occur in more than one clusters, and can not discriminate between clusters.

6 Discussion

In this paper, we try to resolve the relation extraction task in an unsupervised manner. Compared with the existing unsupervised method [7], there are several advantages in our approach.

Relation Types. In [7], each term pair is treated as having one and only one relation type, so they accumulated contexts of all occurrences of a term pair. That is, only one context vector was generated for a term pair. However, our proposed method is based on a more reasonable assumption that there may exist several relation types among different occurrences of a term pair, so, we collect all instances of the occurrences of a term pair, and represent each instance using a context vector. Then our task turns to disambiguate the relation types among the context occurrences of all term pairs.

Context Clustering. [7] adopted a hierarchical clustering method to cluster the contexts. It is very difficult to determine the threshold for the similarity

between clusters, like the appropriate number of clusters. In contrast, through model order selection we can estimate the “nature” number of relation types so that we don’t need to manually pre-define any parameters during the clustering process.

Relation Label. In [7], the author labelled each relation type simply by choosing the most frequent words in a cluster. This method can not ensure that the label of the cluster has the discriminative power since it only considered the frequency of the words. In this paper, we try to discover the feature distribution within a collection, not only in the same cluster but also across the clusters. The features produced by this weighting scheme will provide more valuable characteristics for relation labelling.

Evaluation method. In [7], each cluster is mapped to one ground truth class simply by choosing the one which has the most overlap with it. But two clusters may be mapped to the same relation class, to avoid this bias, we try to find a one to one mapping from estimated cluster to the ground truth classes. Furthermore, we utilize the WordNet to compare the relatedness of our identified labels and the predefined relation labels.

7 Conclusion and Future Work

In this paper, we proposed a method of using model order identification and discriminative feature analysis to improve the unsupervised approach for relation extraction from corpus. The advantages of the proposed approach includes that it doesn’t need any manual labelling of the relation instances, it doesn’t need to pre-define the number of the context clusters, or pre-specify the similarity threshold for the clusters, and it can avoid extracting those common words as characterization of the relations.

One future work is to use some feature selection techniques to acquire good features and incorporate second-order context information to enrich the context information of the entity pairs. Future work also includes using the concepts in Wordnet and other discriminative feature weighting, e.g. information gain, to improve the labelling of the relation types, and applying this technique in ontology construction to acquire the relations between concepts in ontology.

References

1. Defense Advanced Research Projects Agency.: Proceedings of the Sixth Message Understanding Conference (MUC-6). Morgan Kaufmann Publishers, Inc.(1995)
2. Mary Elaine Califf and Raymond J.Mooney.: Relational Learning of Pattern-Match Rules for Information Extraction, AAAI(1999)
3. Sergey Brin.: Extracting patterns and relations from world wide web. In Proc. of WebDB Workshop at 6th International Conference on Extending Database Technology, pages 172-183 (1998)

4. Kiyoshi Sudo, Satoshi Sekine and Ralph Grishman.: An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition. Proceedings of ACL, Sapporo, Japan.(2003)
5. Yangarber,R., R.Grishman, P.Tapanainen, and S.Huttunen.: Unsupervised discovery of scenario-level patterns for information extraction. In proceedings of the Applied Natural Language Processing Conference, Seattle, WA (2000)
6. Eugene Agichtein and Luis Gravano.: Snowball: Extracting Relations from large Plain-Text Collections. In Proc. of the 5th ACM International Conference on Digital Libraries (2000)
7. Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman.: Discovering Relations among Named Entities from Large Corpora. Proceeding of Conference ACL, Barcelona, Spain (2004)
8. Dmitry Zelenko, Chinatsu Aone and Anthony Richardella.: Kernel Methods for Relation Extraction. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia (2002)
9. S.Soderland.: Learning information extraction rules for semi-structured and free text. Machine Learning, 31(1-3):233-272 (1999)
10. Lange,T., Braun,M.,Roth, V., and Buhmann,J.M.: Stability-Based Model Selection. Advances in Neural Information Processing Systems 15 (2002)
11. Levine,E. and Domany,E.: Resampling Method for Unsupervised Estimation of Cluster Calidity. Neural Computation, Vol.13, 2573-2593 (2001)
12. Zhengyu Niu, Donghong Ji and Chew Lim Tan.: Document Clustering Based on Cluster Validation. CIKM'04, Washington, DC, USA, November 8-13 (2004)
13. Volker Roth and Tilman Lange.: Feature Selection in Clustering Problems. NIPS2003 workshop (2003)
14. Gabriel Pui Cheong Fung, Jeffrey Xu Yu and Hongjun Lu.: Discriminative Category Matching: Efficient Text Classification for Huge Document Collections. Proceedings of the IEEE International Conference on Data Mining (ICDM), Maebashi City, Japan, December 09-12 (2002)
15. D.Lin.: Using syntactic dependency as a local context to resolve word sense ambiguity. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, pages 64-71,Madrid, July (1997)
16. P.Resnik.: Using information content to evaluate semantic similarity in a taxonomy. In proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, August (1995)
17. Ted Pedersen, Siddharth Patwardhan and Jason Michelizzi.: WordNet::Similarity-Measuring the Relatedness of Concepts. AAAI (2004)