

# Extracting Relations Within and Across Sentences

**Kumutha Swampillai** and **Mark Stevenson**

Department of Computer Science

Sheffield University

S1 4DP, UK

(k.swampillai|m.stevenson)@dcs.shef.ac.uk

## Abstract

Previous work on relation extraction has focussed on identifying relationships between entities that occur in the same sentence (intra-sentential relations) rather than between entities in different sentences (inter-sentential relations) despite previous research having shown that inter-sentential relations commonly occur in information extraction corpora. This paper describes a SVM-based approach to relation extraction that is applied to both types. Adapted features and techniques for counter-acting bias in SVM models are used to deal with specific issues that arise in the inter-sentential case. It was found that the structured features used for intra-sentential relation extraction can be easily adapted for the inter-sentential case and provides comparable performance.

## 1 Introduction

Relation extraction is an established subfield of information extraction concerned with extracting related pairs of entities from text. The majority of research has been applied to extracting relations within single sentences (intra-sentential relations), examples include (Chieu and Ng, 2002; Culotta and Sorensen, 2004; Sekine, 2006; Banko and Etzioni, 2008). However, an analysis of the MUC6 corpus (Swampillai and Stevenson, 2010) showed that 28.5% of the relations occur between entities in different sentences (inter-sentential relations). This paper describes a SVM-based approach which is applied to the extraction of both inter- and intra-sentential relations.

A number of challenges are faced when extract-

ing inter-sentential relations. The structured features, that are based on parse trees and have been successfully used for intra-sentential relation extraction, do not naturally apply over multiple sentences. The limited research published on inter-sentential relation extraction (Roberts et al., 2008; Hirano et al., 2010) does not employ parse tree features. We address this problem by introducing new structured features (see section 3.2) for the inter-sentential case.

There is also a greater data sparsity issue when learning extraction models for inter-sentential relations due in part to the smaller number of relations expressed inter-sententially. We investigate a learning approach called threshold adjustment (Shanahan and Roma, 2003) to counter-act the imbalance in the data.

The remainder of the paper is organised as follows: Section 2 discusses previous work on relation extraction. Section 3 describes a relation extraction system suitable for both inter- and intra-sentential relation extraction that uses both flat and structures features. The MUC6 relation extraction task is described in Section 4. Section 5 investigates whether the bias in the relation extraction SVM models can be mitigated using threshold adaption. Section 6 reports the results of the inter-sentential and intra-sentential relation extraction system described. Finally, Section 7 concludes the paper with a discussion of the effectiveness of a composite kernel approach to inter-sentential relation extraction.

## 2 Related Work

The majority of the work on relation extraction has focused on intra-sentential relations and there has been limited research on inter-sentential relation extraction. Roberts et al. (2008) applies

an SVM approach to identify inter-sentential relations in the biomedical domain where flat features are used to represent the relations. A low performance is achieved on the inter-sentential relations alone (f-measure  $< 0.19$ ) but they were able to improve overall performance by combining their inter- and intra-sentential data sets.

In addition, Roberts et al. (2008) give a distribution of inter-sentential relations in their corpus where the number of inter-sentential relations occurring in a pair of sentences is inversely proportional to the number of intervening sentences with 42.9% of inter-sentential relations present in consecutive sentences.

More recently Hirano et al. (2010) have reported that 12% of the relations in their Japanese news corpus are inter-sentential. It learns extraction patterns using a bootstrapped classification algorithm. A novel feature is created for inter-sentential relations where a tree is constructed to represent a possible relation based on a *salient referent list*, i.e. a map of the references in the document. The tree contains the two entities and the proposed relation type which is augmented with entity class and POS. An f-measure of 51% is reported for inter-sentential relations.

Flat features commonly used for intra-sentential relation extraction (Mintz et al., 2009) include: a feature representing the entity that occurs first in the sentence; the sequence of lexical tokens and part-of-speech (POS) tags between the two entities, in the sentence; a sequence of lexical tokens and their POS tags on the left hand side of the first entity and on the right hand side of the second entity; a dependency path between the two entities and the verbs that occur between the entities. Composite kernels using flat and structured features have been successfully applied for intra-sentential relation extraction (Zelenko et al., 2003; Bunescu and Mooney, 2004; Culotta and Sorensen, 2004; Zhou et al., 2007). Culotta and Sorensen (2004) and Zhou et al. (2007) have shown that tree kernels combined with flat kernels are more effective for intra-sentential relation extraction than either kernel used alone. In experiments on the ACE corpus, Zhou et al. (2007) achieved f-measures of 0.741 using syntactic parse tree features which outperforms dependency trees. Zhang et al. (2006) further explored which portion of parse trees are most informative for intra-sentential relation extraction by testing seven dif-

ferent subtrees as features. The shortest path-enclosed tree performed the best where the shortest path-enclosed tree is the subtree that includes only the two entities participating in the relation and the intervening syntactic structure.

### 3 Relation Extraction System

We classify relations using SVMs, a standard approach that has been widely used in relation extraction (Agichtein and Gravano, 2000; Zelenko et al., 2003; Roberts et al., 2008; Ittoo and Bouma, 2010). The SVM<sup>light</sup> implementation (Joachims, 2002) and Moschitti's tree kernel tools (Moschitti, 2006) are used. Each pair of entities that appears in the document and is of the correct named entity types is considered a *possible relation* for that relation type. Features are extracted from the text to represent each *possible relations* and these are classified using a binary SVM model. These features are adapted from the set of commonly used features for intra-sentential relation extraction and are based on both flat features and the structured features derived from parse trees. Experiments are also conducted combining the two types of features in composite kernels.

To our knowledge tree and composite kernels have not been applied to inter-sentential relation extraction.

#### 3.1 Flat Features

The entities participating in an inter-sentential relation can occur in any two sentences in a document; therefore the sequence of tokens between the two entities can include a large number of tokens. We therefore use a windowing method to model context of the entities separately. This feature list is given below:

- A window of  $t$  tokens from the surrounding context of each entity.
- A window of  $t$  POS tags from the surrounding context of each entity
- The two nearest dominating verbs for each of the entities, identified in the parse tree structure.
- A distance feature, *dist*, which corresponds to the number of intervening sentences between  $e_1$  and  $e_2$ .

The use of a window to select the token and POS tag features for each entity, instead of the sequence

of tokens between two entities, avoids the situation where document length token sequence is used as a feature. In these experiments two window sizes are used:  $t = 6$  and  $t = 12$  which represent three and six tokens to the left-hand-side and right-hand-side of  $e_1$  and  $e_2$  respectively. The likelihood of a inter-sentential relation is inversely proportional to the distance between the two participating entities and the *dist* feature adds this information to the representation.

### 3.2 Structured Features

Structured features used for intra-sentential relation extraction are based on parse trees. As only entities occurring in the same sentence can be part of a intra-sentential relation, it can be assumed that related entities always appear in a single parse tree. However, this assumption does not hold for inter-sentential relations. We overcame this problem by joining parse trees for pairs of entities by adding a new node (**ROOT**) that connects the parses. Two new features were developed using this approach based on the shortest path-enclosed tree (Zhang et al., 2006):

- The shortest path tree (SPT) structure which only contains the shortest path between the two entities, that is the conjunction of the path from  $e_1$  to the root and the path from  $e_2$  to the root.
- The adapted shortest path-enclosed tree (SPET) consisting of a subtree containing the shortest path between the two participating entities and all intervening nodes and structure to provide context.

Examples are shown in Figure 1.

## 4 Extraction Task

The MUC 6 management succession task identifies information about people entering or leaving management positions in organizations and has been shown to include both inter- and intra-sentential relations (Swampillai and Stevenson, 2010). The main entities participating in these events are the persons joining or leaving (*Per*), the positions they are taking up or vacating (*Post*) and the organizations in which the position exists (*Org*). A version of the MUC6 corpus that has been converted to binary relations is used (Swampillai and Stevenson, 2010), where the

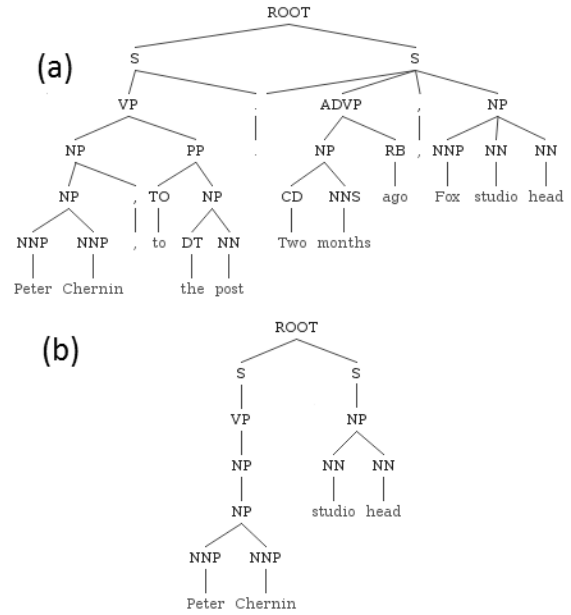


Figure 1: Examples of (a) shortest path-enclosed tree and (b) shortest path tree adapted for inter-sentential relation extraction.

three relation types, *PerOrg*, *PerPost* and *PostOrg*, have been manually identified annotated.

For example, the following sentences include one intra-sentential relation, *PerPost*(*Vern Raburn, president*), and two inter-sentential relations, *PerOrg*(*Vern Raburn, Paul Allen Group*) and *PostOrg*(*president, Paul Allen Group*).

“Paul G. Allen, the billionaire co-founder of Microsoft Corp., has started a company and named longtime friend *Vern Raburn* its *president*.

The company, to be called *Paul Allen Group* will be based in Bellevue, Wash., and will coordinate the overall strategy...”

Intuitively, inter-sentential relation extraction is related to co-reference resolution. However, whilst the resolution of anaphoric expressions can address a significant proportion of these relations, an analysis of the MUC6 corpus by Stevenson (2006) shows that many of these relations require inference across information contained in multiple sentences, possibly using discourse analysis and world knowledge. For example, the following sentences describe a *PerPost* relation where *Kenneth Newell* leaves the position of *senior vice president, Europe, Africa and Mediterranean*.

“David J. Bronczek, vice president and general manager of Federal Express Canada Ltd., was named *senior vice president, Europe, Africa and Mediterranean*, at this air-express concern.

Mr. Bronczek succeeds *Kenneth Newell*, 55, who was named to the new post of senior vice president, retail service operations.”

This relation can only be inferred using the knowledge that when one executive replaces another they must leave the position they are currently holding. This paper proposes an approach that does not require the kind of complex linguistic understanding required for co-reference resolution and addresses all inter-sentential relations.

## 5 Data Sparsity

In the case of intra-sentential relations, possible relations are constrained to pairs of entities that occur within a sentence. Whereas for inter-sentential relations all pairs of entities that occur in a document are possible relations. This causes an explosion in the number of negative instances in the inter-sentential case compared to the intra-sentential case. This coupled with a smaller number of positive relations (only 28.5%) causes a highly unbalanced data set. The percentage of positive examples of relations in all cases is shown in Table 2. It should be noted that there are an extremely limited number of *PerPost* inter-sentential relations, only 64, present in the corpus. This level of imbalance in the data set can render classifiers ineffective (Wu and Chang, 2003).

Relation Type	Intra	Inter
<i>PerOrg</i>	14.99% (1568)	0.53% (29320)
<i>PerPost</i>	23.44% (1971)	0.25% (25697)
<i>PostOrg</i>	20.07% (1495)	0.79% (22475)

Table 1: The bias of the data is expressed here as the percentage of positive relation instances with the total number of instances for each relation type given in brackets.

Various approaches to learning with unbalanced data have been proposed. Undersampling the negative class prior to learning (Japkowicz, 2000) discards a large proportion of the data and the data used for learning no longer approximates the probability distribution of the target population. The

other approach is to introduce a bias in the learning algorithm which compensates for the unbalanced training data without discarding information. Two established methods are cost-sensitive learning (Morik et al., 1999) and hyperplane adjustment (Shanahan and Roma, 2003) both of which have been applied to the relation extraction system. Experiments comparing the two techniques showed that cost-sensitive learning does not perform as well as hyperplane adjustment and these results are not reported here.

### 5.1 Threshold Adjustment

Threshold adjustment is a method for counteracting the bias in SVM models resulting from unbalanced data (Shanahan and Roma, 2003). In the case of unbalanced data the SVM hyperplane is biased towards the negative class, however the hyperplane can be offset so that it preserves the orientation of the original hyperplane but pushes it towards the negative class. The threshold,  $\beta$ , is used to adjust the hyperplane immediately training. Given a set of labelled training instances  $\{x_i, y_i\}_{i=1..n}$  where input points  $x_i$  map to targets  $y_i \in \pm 1$ , the class prediction of a new test instance  $x$  is derived using

$$\text{sign} \left( \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b - \beta \right) \quad (1)$$

where the bias  $b$  and coefficients  $\alpha_i$  are found by SVM training and  $K$  is the kernel function. The constant  $\beta$  is added to bias in the model in favour of the positive instances. Inter-sentential relation extraction is carried out for various values of  $\beta$ , using a prototypical feature selection, including both the flat and structured features. Table 2 gives results for the baseline,  $\beta = 0$ , and the results for the best performing model for each relation type where  $\beta$  maximizes the f-measure. These results show that adjusting the threshold for SVMs can achieve a statistically significant<sup>1</sup> improvement in f-measure over standard SVM models for both relation types.

## 6 Relation Extraction

The relation extraction system described in Section 3 was evaluated on both inter-sentential and intra-sentential relations in the MUC6 corpus. Training and testing was performed using 10-fold nested cross validation.

<sup>1</sup>Statistical significance is tested using the Mann-Whitney U test,  $P < 0.05$ .

Method	PerOrg			PerPost			PostOrg		
	R	P	F-Meas.	R	P	F-Meas.	R	P	F-Meas.
No Bias	0.284	1.000	0.443	0.000	0.000	0.000	0.422	1.000	0.594
Threshold Adaption	0.561	$\beta = -0.75$ 0.920	<b>0.697</b>	0.541	$\beta = -1$ 0.076	<b>0.133</b>	0.668	$\beta = -0.75$ 0.992	<b>0.799</b>

Table 2: Maximum performance boost of cost-sensitive learning and threshold adjustment methods on the performance of inter-sentential relation extraction SVM models.

## 6.1 Nested Cross-Validation

Nested cross-validation (Scheffer, 1999) was used to automatically set the threshold parameter,  $\beta$ , by optimizing it empirically during training. This method also ensures that  $\beta$  is set independently from our testing data. This sub-divides the training data in each cross-validation fold into sub-folds which are used to identify the optimal value of the threshold for that particular training data. This threshold value is then used when evaluating the test data of the original cross-validation fold. The optimal threshold value of each cross-validation fold is identified in the sub-folds by training using a variety of threshold values and evaluating them on the sub-fold reserved for testing. The threshold with the highest average value across all sub-folds is then used. This nested cross-validation algorithm is described in Algorithm 1.

## 6.2 Results

The performance of various feature sets (kernels) is evaluated on both the inter-sentential (Table 3) and intra-sentential (Table 4) relation extraction task. The relation extraction system classifies possible relations from the corpus as one of the three relation types, *PerOrg*, *PerPost* or *PostOrg*. The recall, precision and f-measure metrics is reported after each classifier and kernel. The first three kernels in the tables contain flat features, where *Winn* indicates the inclusion of  $n$  POS tags and tokens surrounding each entity. The *SPT* and *SPET* kernels are the shortest path-enclosed tree and the shortest path tree kernels. The final two are composite kernels combining each tree kernel, *SPT* and *SPET*, with the overall best performing flat kernel, *Win12 + Dist + Verbs*.

The best performance is achieved using the composite SPT kernel for all relation types and for both the inter-sentential and intra-sentential tasks. However, in the case of inter-sentential relations there is no statistically significant difference<sup>2</sup> be-

<sup>2</sup>Statistical significance is tested using the Mann-Whitney U test,  $P < 0.05$ .

---

**Algorithm 1** Procedure for carrying out nested cross-validation to determine the optimal threshold value,  $\beta^*$ , for the training data in each fold. This algorithm extends standard cross-validation by adding an inner loop to estimate the optimal threshold value by finding the maximum f-score for each threshold value,  $\beta$ .

---

```

1: thresholds = {0.25, ..., -1}
2: Split data, T, into 10 folds ( $t_1, t_2, \dots, t_{10}$ )
3: for  $i = 1$  to 10 do
4:   test_set  $\leftarrow t_i$ 
5:   training_set  $\leftarrow T - t_i$ 
6:   Split training_set into 9 folds ( $v_1, v_2, \dots, v_9$ )
7:   for  $j = 1$  to 9 do
8:     testing_validation_set  $\leftarrow v_j$ 
9:     training_validation_set  $\leftarrow$ 
       training_set -  $v_j$ 
10:    Train SVM using the
       training_validation_set, evaluate
       on testing_validation_set and record the
       predictions,  $pred(k)$ .
11:    for all  $\beta \in$  thresholds do
12:      Calculate the f-measure of  $pred(j)$ 
       with a threshold setting of  $\beta$  and
       record,  $F(pred(j))_\beta$ .
13:    end for all
14:  end for
15:  for all  $\beta \in$  thresholds do
16:     $F_{avg}(\beta) \leftarrow \frac{\sum_{j=1}^{10} F(pred(j))_\beta}{9}$ .
17:  end for all
18:  Determine the best threshold setting,  $\beta^*$ ,
       where  $\beta^* = \text{argmax } F_{avg}(\beta)$ .
19:  Train the SVM using training_set, evalu-
       ate on test_set with  $\beta^*$  as threshold setting
       and record performance,  $P(i)$ 
20: end for
21: performance  $\leftarrow \frac{\sum_{i=1}^{10} P(i)}{10}$ 
22: return performance

```

---

tween the performance of the SPT kernel and the composite SPT kernel on both *PerOrg* and *Pos-*

Kernel	<i>PerOrg</i>			<i>PerPost</i>			<i>PostOrg</i>		
	R	P	F-Meas.	R	P	F-Meas.	R	P	F-Meas.
<b>Flat</b>									
Win 6+Dist	0.117	0.730	0.201	0.015	0.200	0.029	0.336	0.809	0.475
Win 12+Dist	0.191	0.644	0.295	0.075	0.440	0.128	0.400	0.681	0.504
Win 12+Dist+Verbs	0.517	0.740	0.608	0.059	0.500	0.106	0.677	0.743	0.708
<b>Tree</b>									
SPT	0.467	0.798	0.589	0.000	0.000	0.000	0.524	0.814	0.638
SPET	0.314	0.608	0.414	0.035	0.167	0.058	0.475	0.656	0.551
<b>Composite</b>									
SPT and Win 12+Dist+Verbs	0.518	0.877	<b>0.651</b>	0.144	0.327	<b>0.200</b>	0.693	0.853	<b>0.765</b>
SPET and Win 12+Dist+Verbs	0.442	0.762	0.560	0.072	0.300	0.116	0.588	0.777	0.669

Table 3: Performance of *inter-sentential* relation extraction for flat, tree and composite kernels using threshold optimization.

Kernel	<i>PerOrg</i>			<i>PerPost</i>			<i>PostOrg</i>		
	R	P	F-Meas.	R	P	F-Meas.	R	P	F-Meas.
<b>Flat</b>									
Win 6	0.535	0.484	0.508	0.645	0.588	0.615	0.614	0.550	0.581
Win 12	0.628	0.441	0.519	0.654	0.561	0.604	0.521	0.503	0.512
Win 12+Verbs	0.589	0.459	0.516	0.660	0.571	0.612	0.415	0.596	0.489
<b>Tree</b>									
SPT	0.566	0.636	0.599	0.630	0.631	0.631	0.623	0.754	0.683
SPET	0.616	0.414	0.495	0.576	0.575	0.575	0.564	0.538	0.551
<b>Composite</b>									
SPT and Win 12+Verbs	0.757	0.649	<b>0.699</b>	0.682	0.624	<b>0.652</b>	0.759	0.741	<b>0.750</b>
SPET and Win 12+Verbs	0.568	0.560	0.564	0.595	0.628	0.611	0.685	0.668	0.677

Table 4: Performance of *intra-sentential* relation extraction for flat, tree and composite kernels using threshold optimization.

*tOrg* relations. This shows the minimal contribution of flat features to the inter-sentential classification task, unlike the intra-sentential task where the addition of flat features makes a marked improvement.

For both tasks the relation type with the best f-measure is *PostOrg* at 0.809 and 0.750 for the inter- and intra-sentential relations respectively. The data set associated with this relation is the least skewed of the data sets. In contrast *PerPost*, the most unbalanced data set, has the worst f-measure for the intra-sentential relation extraction task at 0.652 and fails to make any impact on the inter-sentential relation extraction task with an f-measure of only 0.200. This suggests that bias still has an effect on performance despite the steps taken to mitigate against it.

Different behaviour is observed for inter- and intra-sentential relations when comparing the results of the experiments using the flat kernel. The use of a wider context feature window and surrounding verbs improves the overall f-measure scores for inter-sentential relations, substantially improving recall while slightly degrading precision. However, for the intra-sentential case adding

context and verb features either maintains or degrades performance. Flat features alone achieve better performance for the inter-sentential task (0.608, 0.128 and 0.708) than for intra-sentential task (0.519, 0.615 and 0.581).

Results using tree and composite kernels show that the SPT tree representation is more effective than the SPET tree for both tasks. This may be because SPET subtrees are larger and potentially contain more noise. Tree kernels perform better than those created from flat features demonstrating that structured features are hugely informative for relation extraction.

Overall, the results show that the best performing kernel is the composite SPT kernel. This is inline with previous research into intra-sentential relation extraction (Zhou et al., 2007; Zhang et al., 2006) where the best results are achieved with a shortest path composite kernel. For inter-sentential relations f-measures of 0.651, 0.200 and 0.809 are achieved. The use of the composite kernel SVM approach to relation extraction gives comparable performance on the inter-sentential task except in the case of relations with extremely skewed training data.

## 7 Conclusions

This paper investigates whether state-of-the-art approaches to intra-sentential relation extraction can be effectively adapted for inter-sentential relation extraction. The results demonstrate that a composite kernel approach to inter-sentential relation extraction can achieve comparable results with intra-sentential relation extraction. We have also shown that the structured features used for intra-sentential relation extraction can be easily adapted for the inter-sentential case. The performance of structured features has been found to be superior to flat features which have previously been used for the inter-sentential relation extraction task (McDonald et al., 2005; Roberts et al., 2008).

Overall, composite kernels, that combine a larger context window with a SPT tree, were found to give better performance than either flat or structured features alone. Inter-sentential *PerPost* relations could not effectively be extracted using this approach, most likely due to the bias in the *PerPost* data set.

Threshold adaption, which was optimised using nested cross-validation, significantly improved the performance of SVM models for inter-sentential relation extraction. Average f-measure improved from 0.295 to 0.605, a significant improvement in performance over all kernel types.

## References

- Eugene Agichtein and Luis Gravano. 2000. Snowball: Extracting Relations from Large Plain-Text Collections. *In Proceedings of the 5th Association for Computing Machinery International Conference on Digital Libraries*, 85–94.
- Michele Banko and Oren Etzioni. 2008. The Tradeoffs between Open and Traditional Relation Extraction. *In proceedings of the 46th Association for Computational Linguistics Annual Meeting and Human Language Technology Conference (ACL-08:HLT)*, 28–36.
- Razvan Bunescu and Raymond J. Mooney. 2004. Collective Information Extraction with Relational Markov Networks. *In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL 2004)*, 438–446.
- Hai Leong Chieu and Hwee Tou Ng. 2002. A Maximum Entropy Approach to Information Extraction from Semi-Structured and Free Text. *In Proceedings of the Eighteenth International Conference on Artificial Intelligence (AAAI02)*, 768–791.
- Aron Culotta and Jeffrey Sorensen. 2004. Dependency Tree Kernels for Relation Extraction. *In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL 2004)*, 423–430.
- Toru Hirano, Hisako Asano, Yoshihiro Matsuo and Genichiro Kikui. 2010. Recognizing Relation Expression between Named Entities based on Inherent and Context-dependent Features of Relational words. *In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010: Posters)*, 409–417.
- Ashwin Ittoo and Gosse Bouma. 2010. On Learning Subtypes of the Part-Whole Relation: Do Not Mix Your Seeds. *In Proceedings of the 48th Annual Meeting on Association for Computational Linguistics (ACL 2010)*, 1328–1336.
- Nathalie Japkowicz. 2000. The Class Imbalance Problem: Significance and Strategies. *In Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI)*, 111–117.
- Thorsten Joachims. 2002. *Learning to Classify Text using Support Vector Machines*. Kluwer Academic Publishers, Norwell, MA, USA.
- Ryan McDonald, Fernando Pereira, Seth Kulick, Scott Winters, Yang Jin, and Pete White. 2005. Simple Algorithms for Complex Relation Extraction with Applications to Biomedical IE. *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL '05)*, 491–498.
- Tara McIntosh and James R. Curran. 2007. Challenges for Extracting Biomedical Knowledge from Full Text. *In Proceedings of the Workshop on Biological, Translational and Clinical Language Processing (BioNLP 2007)*, 171–178.
- Mike Mintz, Steven Bills, Rion Snow and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. *In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 1003–1011.
- Katharina Morik, Peter Brockhausen and Thorsten Joachims. 1999. Combining Statistical Learning with a Knowledge-Based Approach - A Case Study in Intensive Care Monitoring. *Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99)*, 268–277.
- Alessandro Moschitti. 2006. Making Tree Kernels Practical for Natural Language Learning. *In Proceedings of the 11st Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, 113–120.
- Angus Roberts, Robert Gaizauskas and Mark Hepple. 2008. Extracting Clinical Relationships from Patient Narratives. *In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, 10–18.

- Tobias Scheffer. 1999. Error Estimation and Model Selection. *Technischen Universität Berlin, School of Computer Science*.
- Satoshi Sekine. 2006. On-Demand Information Extraction. *In Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, 731–738.
- James G. Shanahan and Norbert Roma. 2003. Boosting Support Vector Machines for Text Classification through Parameter-free Threshold Relaxation. *In Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM '03)*, 247–245.
- Mark Stevenson. 2006. Fact Distribution in Information Extraction. *Journal of Language Resources and Evaluation*, 40:183–201.
- Kumutha Swampillai and Mark Stevenson. 2010. Inter-sentential Relations in Information Extraction Corpora. *In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010)*, 2637–2641.
- Vladimir N. Vapnik. 1995. *The nature of statistical learning theory*. Springer-Verlag Inc., New York, USA.
- Konstantinos Veropoulos, Colin Campbell and Nello Cristianini. 1999. Controlling the Sensitivity of Support Vector Machines. *Proceedings of the International Joint Conference on AI (IJCAI 1999)*, 55–60.
- Gang Wu and Edward Y. Chang. 2003. Class-Boundary Alignment for Imbalances Dataset Learning. *In ICML 2003 Workshop on Learning from Imbalanced Datasets II, Washington, DC*.
- Dmitry Zelenko, Chinatsu Aone and Anthony Richardella. 2003. Kernel methods for Relation Extraction. *Journal of Machine Learning Research*, 3:1083–1106.
- Min Zhang, Jie Zhang, Jian Su and Guodong Zhou. 2006. A Composite Kernel to Extract Relations between Entities with both Flat and Structured Features. *In Proceedings of the 21st International Conference on Computational Linguistics (Coling) and the 44th annual meeting of the Association for Computational Linguistics (ACL)*, 825–832.
- GuoDong Zhou, Min Zhang, DongHong Ji and QiaoMing Zhu. 2007. Tree Kernel-Based Relation Extraction with Context-Sensitive Structured Parse Tree Information. *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 728–736.