
Learning Relations from Biomedical Corpora Using Dependency Tree Levels

Sophia Katrenko

Human-Computer Studies Laboratory, University of Amsterdam, Kruislaan 419, 1098VA, Amsterdam

KATRENKO@SCIENCE.UVA.NL

Pieter Adriaans

Human-Computer Studies Laboratory, University of Amsterdam, Kruislaan 419, 1098VA, Amsterdam

PIETERA@SCIENCE.UVA.NL

Abstract

In this paper we address the problem of learning relations in the biomedical domain. We propose a representation which takes into account the syntactic information and allows for using different machine learning methods. The results we have obtained are comparable to the performance of relation learning systems in the biomedical domain and in some cases out-perform them. In addition, we have studied the impact of ensemble methods on learning relations using the representation we proposed. Given that recall is very important for relation learning, we have studied how it can be improved. It has been shown that ensemble methods techniques provide recall of 76,2% (with precision of 69,4%).

1. Introduction

Not only the number of publications in the biomedical domain grows rapidly every year, there are also many approaches proposed to how to handle such amount of data.

These approaches primarily consider such tasks as text mining, information extraction and information retrieval. Information retrieval focuses on the retrieval of the full documents, while the goal of information extraction is to find text fragments relevant to the user need. On the other hand, it is often useful to get more fine-grained information, for instance, the list of biomedical instances or relations. Such information might be especially important for the curation of existing resources, such as databases of interactions (Albert et al., 2003).

The paper is organized as follows. We first start with a discussion of related work and a problem statement.

In the section 3, we present our approach and provide motivation for it. Further, we test our approach on two data sets for interaction extraction. We report on our results and conclude with the discussion and outlook for future work.

2. Problem Statement and Related Work

The biggest collection of the medical documents is Medline with its 2,000 citations added every week. The large size of this collection makes it impossible to annotate it all by humans. Consequently, there have been several attempts to create smaller annotated corpora based on Medline, such as Genetag used for the gene/protein named entity recognition (NER) (Tanabe et al., 2005), or MedTag, the corpus comprising Genetag, MedPost and ABGene (Smith et al., 2005). There have also been corpora created with a special purpose to be used by the various challenges, e.g. corpus of the annotated gene-protein relations for the "Genic Interaction Extraction Challenge" (Nédellec, 2005).

In general, the relation learning problem can be seen as a two-step process. First, the relation arguments have to be identified. Further, it is necessary to check whether the relation holds. This setting has also been used for the relation discovery in other domains (Zelenko et al., 2003), moreover, it is often assumed that the arguments have already been found. In this case, the relation learning is reduced to the second step which involves procedures enabling such verification. It has been shown by Bunescu et al. (2005) that provided the correct names of proteins are given, the accuracy of relation discovery is much higher. An interesting observation has been made by Cohen and Hersh (2005) who considered binary biomedical relations. Although the accuracy of relation extraction for many domains (such as news article extraction) cru-

cially depends on the accuracy of named entity recognition and is equal to the cube¹ of the performance of latter, it seems not to hold in the biomedical domain. The conclusion can be drawn that the surrounding context makes it easier to identify the arguments of a relation in the biomedical domain.

The relation learning task can be formulated in the following way:

Definition 1 (Relation learning) *Given a data set D ² and an n -ary relation Rel with the arguments $X, Y \dots Z$ find all instances $x \in X, y \in Y, \dots, z \in Z$ ($x, y, z \in D$), such that $Rel(x, y, \dots, z)$ holds.*

Below, we discuss the relation learning task from the following perspectives: types of relations and approaches.

Relation learning has received much attention in the past decade but it is nevertheless difficult to compare the results obtained by different research groups. It concerns not only the data sets being used but also the types of relations in question. Often, a certain relation is in focus, such as inhibition (Pustejovsky et al., 2002) or a relation between genes and diseases. The latter is a causal relation which can be formulated as a question "Which gene(s) cause(s) a disease Y?" which in turn can be used for question answering or information retrieval (Hersh & et al., 2005). This type of relation has been studied by Craven and Kumlien (1999), Ray and Craven (2001). More recently, there has been work on the gene-disease relations carried out by Chun et al. (2006). Contrary to the approach taken by Craven and Kumlien (1999) who have used weakly labeled data, Chun et al. aimed at using a corpus annotated by humans. However, it has been shown that if a gene and a disease co-occur, they are likely to be true positives for the relation extraction. 94% of the correctly identified and co-occurring genes and diseases presented a gene-disease relation. We assume, therefore, that in the discovery of a gene-disease relation, it is necessary to study recall.

However, it is necessary to achieve high recall in other relation learning tasks as well. Some of the relations, such as interactions between genes or genes and proteins are more complex. They can be further divided into groups according to type of interaction, such as

¹Since a relation in question is binary, one needs to identify two arguments and a term identifying a relation itself. It is therefore assumed that the performance of the relation identification equals to the cube of the performance of identifying each of the arguments and a link-word

²where a data set D can be text, semi-structured data, etc.

interactions expressed by explicit action, binding of the protein on the promoter of a gene, etc. Since the arguments of such relations are genes or proteins, it is important to know whether a given relation is symmetric. The asymmetry of a relation also rises the rate of false positives.

Most approaches to relation learning fall in one of two categories, either *hand-written* patterns or *learning oriented* approaches. The approaches based on the hand-written (mostly pattern-oriented) are usually time-consuming since they often assume use of rules (patterns) written by an expert. Consequently, when such rules are applied to unseen data, they fail to take into account relations expressed in another way. Although patterns provide a high precision, recall might be much lower (Thomas et al., 2000). In the biomedical domain, it has been proposed to use two types of patterns. The first type is sequential and based on the often occurring sequences of words in a sentence. The second type (Khoo et al., 2000) attempts to account for a syntactic structure of a sentence. Taken the dependency structure of a sentence, which is usually represented as a tree, the patterns in the latter case are subtrees. Such patterns are sometimes referred to as graphical (Khoo et al., 2000). A simpler approach is to consider not the dependency tree as a whole but certain predefined syntactic functions. This idea has been used by Hahn and Romacker (2000) and Rinaldi et al. (2004) who have focused on the *subject-verb-object patterns*.

The drawback of the approaches using hand-written patterns is their low recall. Another way to construct such patterns is to use a rule learning algorithm. The performance of the rule learning methods has been studied in detail by Bunescu et al. (2005). The authors have addressed the problems of protein identification and extraction of the protein interactions. For the relation extraction, two approaches have been developed, based on the rule learning method Rapiere and on the longest common subsequences. It has been shown that these two approaches outperform the hand-written rules.

Contrary to the approaches discussed above, Pustejovsky et al. (2002) and Leroy and Chen (2005) have employed finite state automata to learn relations. When testing their approach on the inhibit-relation, Pustejovsky et al. (2002) have received precision of 90,4% and recall of 58,9%. A particularly interesting approach has been proposed by Bunescu and Mooney (2005) who have studied subsequence kernels for relation extraction. Comparative experiments on the AImed corpus (Bunescu & Mooney, 2005) have re-

vealed that the relation kernel outperforms the approaches based on the longest common subsequences and hand-written rules.

Approaches based on a pure co-occurrence of the biomedical terms are also useful, however, their performance depends on the type of a relation (Stephens et al., 2001). As mentioned above, the co-occurrence of terms denoting diseases and genes is likely to provide evidence for the relation between them. In contrast to a gene-disease relation, a relation between genes is less predictable by the pure co-occurrence of genes in a sentence.

Since there are many knowledge resources created in the biomedical community over past years, it is especially valuable to test their impact on the biomedical entity extraction task. Leroy and Chen (2005) have presented a hybrid system integrating linguistic parsing with the existing knowledge sources, such as Gene Ontology, UMLS, and the HUGO nomenclature. They have evaluated 549 relations from Medline abstracts containing p53 gene. In comparison to the relations extracted by parser, the relations provided by a co-occurrence based semantic net Concept Space have been less precise and relevant. However, when adding relations containing terms found in GO and HUGO, precision increases. By such approaches as Leroy’s one, it has been demonstrated that the knowledge sources can contribute to the protein identification or relation extraction tasks.

3. Approach

Our approach makes use of the syntactic information, though in a different way than the methods described in (Khoo et al., 2000). In what follows, we present our method and give a motivation from both, linguistic and machine learning perspectives.

The approach we take follows the definition of relation discovery as a two-step process, concentrating on the second step only. We assume that we have already identified the arguments of a relation.

In the grammatical tradition, a syntactic structure of a sentence can be presented either by concordance relations or by dependency relations (Rastier et al., 2001). Dependency is a hierarchical relation where every word in a sentence is linked to a word dominating it. Graphically, such dependency relations are presented as a tree, moreover, a root of a tree is often a verb. As shown by Sekimizu et al. (1998), a relation between two or more arguments is also often expressed by verbs. We can therefore conclude that a root of a dependency tree conveys information crucial for the relation learn-

ing. Furthermore, by examining a parent and children of a given node, one can notice that they constitute a local context important for the argument identification. The closest approach to ours has been taken by Hacioglu (2004) who has used dependency trees, in particular, information from the tree levels, for the semantic role labeling. However, the task of semantic labeling is different from the relation learning.

The dependency structure for the sentence (1) is presented on Fig.1. This sentence is part of the AImed data set. In it, *Cdc25* and *Raf1* are interacting proteins. The root of the dependency tree depicted on Fig.1 consists of the word *activated*.

- (1) *Cdc25* can be activated in vitro in a *Raf1*-dependent manner.

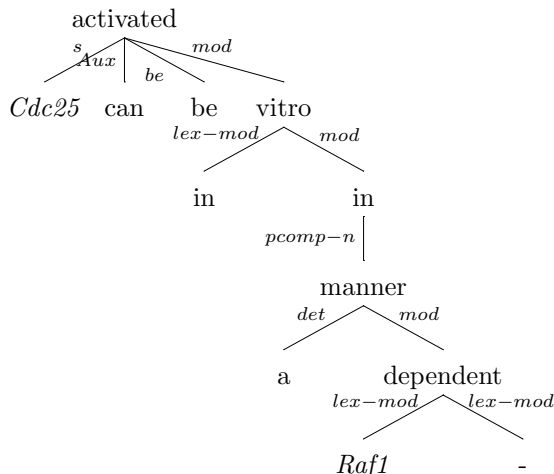


Figure 1. Dependency structure

There are several advantages in considering the dependency tree levels. First of all, it is possible to test our hypothesis in order to discover which levels are the most important for the relation learning. Selecting tree levels can be considered as a feature engineering step. Moreover, since the final representation is of the attribute-value type, it is possible to test different machine learning methods. It is of considerable interest to apply ensemble methods (Dietterich, 2000). The experiments for the named entity recognition task have already demonstrated that use of meta-learning improves the accuracy of classification (Sang & Meulder, 2003).

As already mentioned, we divide all features into two groups, local and global context. To reduce data sparseness, we decided to use lemmata³ instead of

³Lemmata are canonical forms of lexemes. For nouns

words. A parent of a given node (P) and its two children form a local context. The features of a parent and a child are lemmata and the syntactic function between a node in question and a parent (a child). Since a tree is an acyclic graph, each node has at most one parent but can have more than one child. We limited ourselves to two children, C^1 and C^2 .

A global context consists of a least common subsumer (LCS) and a root of a tree (R).

Definition 2 (Least common subsumer (LCS))

Given two nodes A and B in a dependency tree T , a least common subsumer $LCS(A, B)$ is a node L , such that L is ancestor for both, A and B , and there exist no other node N being an ancestor for A and B , such that L is ancestor of N . There is exactly one LCS for any two nodes in a dependency tree.

For example, for words a and $Raf1$ on Fig.1, the least common subsumer is *manner*. Although such nodes, as *in*, *vitro*, and *activated* are all ancestors of a and $Raf1$, they are not least common subsumers.

Some parsers treat a subordinate clause as separate producing not a single tree for a sentence but two. We decided therefore to define two features, one for a root of the first argument (R^1) and a second for a root of the second argument (R^2). Table 1 illustrates how the features have been grouped into feature sets given two nodes X and Y and the relation $Rel(X, Y)$.⁴

Table 1. Feature sets

Feature set(FS)	Features
FS1	LCS
FS2	$C_X^1, C_X^2, C_Y^1, C_Y^2, LCS$
FS3	P_X, P_Y, LCS
FS4	C_X^1, C_Y^1, P_X, P_Y LCS, R_X
FS5	$C_X^1, C_X^2, C_Y^1, C_Y^2, P_X, P_Y$ LCS, R_X, R_Y

When considering the fifth data set (FS5), the example on Fig.1 can be represented as on Fig.2.

Note that we incorporated the syntactic labels into the parent-features P . $Cdc25$ is linked to the word *activated* by the syntactic function s (standing for a

they usually are nouns in the singular, nominative case (such as lemma *dog* for a word *dogs*), for verbs lemmata represent verbs in the infinitive (e.g., the lemma *go* for the word *went*)

⁴In Table 1, the lower indices correspond to two arguments, X and Y

subject), while $Raf1$ is connected to *dependent* by *lexmod* function (standing for a modifier).

Table 2. Feature set for the example on Fig.1

Feature	Cdc25	Raf1
C^1	-	-
C^2	-	-
P	activate_s	dependent_lexmod
LCS	activate	activate
R	activate	activate

4. Experiments

4.1. Data sets

For our experiments, we have used two data sets. One of them is AImed (Bunescu et al., 2005) and the other is a data set created within the "Genic Interaction Extraction" challenge (Nédellec, 2005) (from now, we refer to it as LLL (Learning Language in Logic) data set). The AImed data set consists of the examples of protein-protein interactions. It has been compiled from the 225 Medline abstracts and annotated by the experts. The second data set, LLL, has been created by extracting Medline abstracts on *Bacillus subtilis*. It also includes annotations created by experts, with the distinction that the focus is on the interactions between genes and proteins. The LLL data set consists of 77 sentences and 161 annotated interactions.

4.2. Data Preprocessing

The LLL data set already consists of the tokenized sentences accompanied by the syntactic analysis. For parsing, the LLL organizers have used LinkParser whose output has been verified by experts. Besides this, the dictionary of genes and proteins have been provided so we annotated all occurrences of the dictionary items in the text as biological entities (in the dictionary, no distinction between genes and proteins has been made).

The second data set has been preprocessed by us. The preprocessing steps included tokenization and parsing. We have used a tokenizer based on the white spaces and the parser Minipar. Minipar is a dependency parser whose precision equals to 88% and whose recall equals to 80% on the Susanne corpus⁵. We have also found that the present annotation sometimes contains protein tags surrounding the interaction tags as shown

⁵Minipar is available from <http://www.cs.ualberta.ca/~lindek/minipar.htm>

in (2). Here, *Ras* is annotated as a protein being an argument of an interaction with *RIN1*. In addition, *Ras binding protein* is also annotated as a protein. As explained below, we have constructed false interactions for the training purpose based on the entities annotated as proteins and not being part of a relation. We scanned AImed data set and found 14 cases where annotation of a protein included annotation of interactions as its part. While carrying out preprocessing, the external protein tags have been removed.

- (2) Human <p1 pair="1"><prot>RIN1</prot></p1> was first characterized as a <prot>p2 pair="1"><prot>Ras</prot></p2> binding protein </prot> based on the properties of its carboxyl-terminal domain.

The data sets we have used provide annotations of the binary interaction relation. In order to obtain the negative interactions, we have followed the closed world assumption. However, it has been used in a different way for each of the data sets. The interactions between proteins in AImed are considered to be symmetric. Therefore, the false positives are created as all pairs of proteins being not arguments of the interaction relation as well as pairs where one of the protein is an argument of a relation in a given sentence. LLL data set contains interactions between genes and proteins which are treated as instances of an asymmetric relation. Because of this, the false interactions are produced as pairs of biomolecular entities (i.e., proteins or genes) which do not participate in a relation but also those where the arguments of a relation are flipped (e.g., a pair (X,Y) where X and Y are biomolecular entities will be considered a false positive for the true interaction (Y,X)).

After constructing a training set, we received 873 training instances for the LLL data set, 161 of which were positive examples. For the AImed corpus, we obtained 4,736 instances with 985 of them being positive examples.

5. Results and Discussion

The results we present below have been received by 10-fold cross-validation for AImed data set and 5-fold cross-validation for the LLL data set, respectively. We have also used the implementation of the machine learning methods from the Weka toolkit (Witten & Frank, 2005).

We studied the impact of the feature sets mentioned above on recall and precision. First, we started with a feature set containing the least common subsumer

only (FS1). In our view, this feature set is similar to the pattern approaches whose main objective is to find a link (a so-called relation word) between two arguments.

Table 3. Least common subsumer: AImed data set

WORDS (LCS)	OCCURRENCE
OF	434
BE	341
PROTEIN	339
...	
bind	139
interact	134
COMPLEX	59
inhibit	52
SHOW	49
DETECT	37
ASSOCIATE WITH	36
REVEAL	35
CONTAIN	34
induce	33
activate	31
REQUIRE	27
EXPRESS	27
ASSOCIATE	26
regulate	25
suppress	14
...	

As the Table 3 suggests, the least common subsumer often includes the words important for relation learning in the biomedical domain. In the list above, such words are bold-faced. Comparing this list to list of verbs identifying relations (Sekimizu et al., 1998), we found that they significantly overlap. However, the

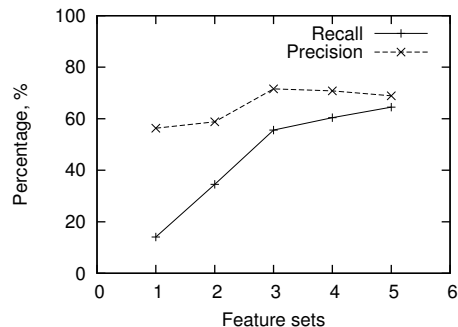


Figure 2. Precision and recall for different feature sets (BayesNet classifier, AImed data set)

use of this feature set results in a very low recall. This supports our hypothesis about a precompiled list of patterns used for relation extraction - in most cases, they cannot cover unseen data very well. Our results

are in line with those reported by Ahmed et al. (2005).

The second feature set, FS2, consists of FS1 and the children of two arguments. As the results on Fig.2 suggest, recall can be already improved by adding the information about the children.

However, it is not satisfactory. Using the third data set containing lemmas from the parent-level provides much better results. We believe it is due to the fact that in many cases proteins are leaves in a tree so the information about the children is missing. The best performance has been obtained by employing the fifth feature set containing all features as defined in Section 3. The precision-recall curve for this experiment is presented on Fig.3.

As we have mentioned above, in most cases the existing approaches to relation extraction provide considerably high precision but low recall. Our results suggest that the approach we have taken also leads to higher precision and usually lower recall. It can be concluded that although the information from the dependency tree levels helps to find many true positives, the local context is sometimes not sufficient to be able to discriminate between true and false positives.

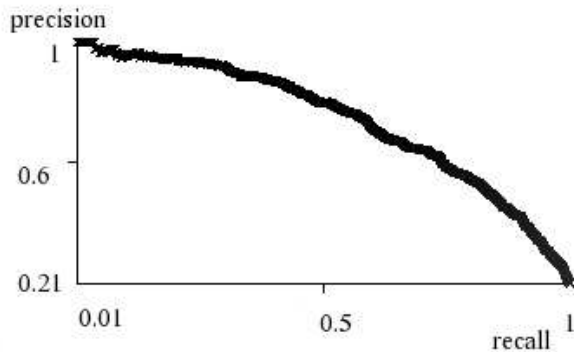


Figure 3. Precision-recall curve for the F5 feature set (BayesNet classifier)

To test the hypothesis that ensemble methods may improve the overall performance, we have conducted preliminary experiments with three ensemble methods, stacking, bagging and AdaBoost. Ensemble of classifiers provide better accuracy if the individual classifiers are diverse and accurate. A classifier is said to be accurate when its error rate is better than random guessing (Dietterich, 2000). Bagging and AdaBoost present the ensemble methods manipulating the training examples. The main idea behind such methods lies in generating multiple hypotheses. In case of bagging,

a different subset from the training data is sampled every time a learning algorithm is applied. Bagging and AdaBoost work well for such algorithms as rule learning or decision tree methods which are generally considered to be unstable. Stacking belongs to the method of combining different classification models.

Table 4. Results on AImed data set

METHOD	PRECISION	RECALL	F-SCORE
NAÏVE BAYES	71,5%	57,6%	63,8%
BAYESNET	68,9%	64,5%	66,6%
IB3	81,3%	51,6%	63,1%
IB1	77,4%	66,3%	71,4%
STACKING	69,4%	76,2%	72,7%
BAGGING	68,2%	63,7%	65,8%
ADABOOSTM1	67%	68,7%	67,8%

We considered three classifiers of a different nature, BayesNet, Naïve Bayes method and K-nearest neighbor classifier (IB1, IB3). Bagging and AdaBoost have been applied with BayesNet classifier. The experiment with stacking has been constructed in the following way: BayesNet has been chosen as meta-classifier with NaïveBayes and 1-nearest neighbour classifiers as individual classifiers. As shown on Table. 4, stacking has proven to provide much higher recall.

(Bunescu & Mooney, 2005) have also used 10-fold cross-validation to test their methods on the AImed corpus. They reported on the performance of their approaches by presenting it as a precision-recall curve. To compare our results on the AImed corpus with the performance of the methods described in (Bunescu & Mooney, 2005), we chose the highest recall received by stacking (76,2%). It corresponds to the precision of 40% on the precision-recall curve which has been received by the subsequence kernel method. We can conclude therefore that our method outperforms the subsequence kernel method on AImed corpus by 29,4%. The comparison of the results of the best individual classifier (BayesNet) to the subsequence kernel method also demonstrates that the first performs better than the latter. The difference in the performance can be explained by the features used in (Bunescu & Mooney, 2005). In particular, Bunescu and Mooney (2005) have considered sequential information which consisted of the words found between two entities, in front of them and after them. After having defined such features, the authors have restricted themselves to the subsequences of the types mentioned above, where a maximum word length equals 4. According to Bunescu and Mooney (2005), such feature selection leads to less overfitting. In our case, we considered not the common

subsequences but the levels from the dependency tree instead. We believe that the selection of levels we have made provides information sufficient for the relation learning and constitutes an alternative approach to the method proposed by Bunescu and Mooney (2005).

Table 5. Results on LLL data set

METHOD	PRECISION	RECALL	F-SCORE
NAÏVE BAYES	56,7%	34,2%	42,6%
BAYESNET	64,7%	53,4%	58,5%
IB1	65,8%	32,3%	43,3%

Comparison of the results received on the AImed data set with the performance on the LLL data set (Table 5) demonstrates that we have received much better results on the first corpus. There are several distinctions between two data sets. First, LLL data set is much smaller, and, second, the underlying assumption behind LLL data set is that all annotated relations are asymmetric. In contrast to this, it is assumed that AImed data set presents symmetric relations. Some classification errors on the LLL data set can be explained by the asymmetry of the relation between genes and proteins.

In general approaches making use of the syntactic structure depend on the accuracy of the parser. For example, precision of the parser we used is 88% and it is likely that some errors in classification are due to the incorrect parsing. While comparing the results on AImed data set and on the LLL data set, we also discovered that the output of Minipar and LinkParser, which has been used by the organizers of LLL, differs. This also affects classification accuracy. For instance, the phrases, such as *activation in prespore* are treated differently by Minipar and LinkParser. While the first introduces a special link *pcomp-n* between a preposition *in* and a noun *prespore*, the latter incorporates preposition in the syntactic function and outputs the relation *comp-in* between *activation* and *prespore*. The performance of current state-of-art parsers on the biomedical data has been studied by Grover et al. (2005). The evaluation of a parser has also been done by Rinaldi et al. (2004) who used LT Chunk to obtain verbal and nominal chunks. Nevertheless, the results Rinaldi et al. (2004) have achieved with the correct (verified) syntactic analysis do not differ much from the results received by the parser.

6. Conclusions

In this paper, we have proposed a representation for learning relations based on the dependency trees. We

have tested this representation on the data sets containing interactions between genes and proteins (LLL data set) and between proteins (AImed data set). The results on the AImed data set are promising and better than the results reported by Bunescu and Mooney (2005). One of the directions in our future research is to carry out a deeper comparison between our approach and the method proposed by Bunescu and Mooney (2005). Since the feature sets in both approaches are different, we plan to explore whether these two methods can complement each other. We also conducted preliminary experiments using ensemble methods whose main purpose is to combine the decisions of individual classifiers. The performance of some of the ensemble methods, such as stacking, suggest that they provide higher recall. We plan to investigate the ensemble methods further in our future research.

Acknowledgment

This work was carried out in the context of the Virtual Laboratory for e-Science project (www.vl-e.nl). This project is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

References

- Ahmed, S. T., Chidambaram, D., Davulcu, H., & Baral, C. (2005). Intex: A syntactic role driven protein-protein interaction extractor for bio-medical text. *In Proceedings of the ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics* (pp. 54–61). Detroit: Association for Computational Linguistics.
- Albert, S., Gaudan, S., Knigge, H., Raetsch, A., Delgado, A., & Huhse, B. (2003). Computer-assisted generation of a protein-interaction database for nuclear receptors. *Molecular Endocrinology*.
- Bunescu, R. C., Ge, R., & Kate, R. J. (2005). Comparative experiments on learning information extractors for proteins and their interactions. *Artificial Intelligence in medicine*, 33, 139–155.
- Bunescu, R. C., & Mooney, R. J. (2005). Subsequence kernels for relation extraction. *In Proceedings of the 19th Conference on Neural Information Processing Systems*.
- Chun, H. W., Tsuruoka, Y., Kim, J. D., Shiba, R., & Nagata, N. (2006). Extraction of gene-disease re-

- lations from medline using domain dictionaries and machine learning. *In Proceedings of the 11th Pacific Symposium on Biocomputing*.
- Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1), 57–71.
- Craven, M., & Kumlien, J. (1999). Constructing biological knowledge bases by extracting information from text sources. *In Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology* (pp. 77–86). Heidelberg, Germany: AAAI Press.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science*, 1857, 1–15.
- Grover, C., Lascarides, A., & Lapata, M. (2005). A comparison of parsing technologies for the biomedical domain. *Natural Language Engineering*, 11(1), 27–65.
- Haciouglu, K. (2004). Semantic role labeling using dependency trees. *In Proceedings of CoLING-04*. Geneva, Switzerland.
- Hahn, U., & Romacker, M. (2000). An integrated model of semantic and conceptual interpretation from dependency structures. *Proceedings of the 18th conference on Computational linguistics* (pp. 271–277). Morristown, NJ, USA: Association for Computational Linguistics.
- Hersh, W., & et al. (2005). Trec 2005 genomics track overview. *TREC 2005 meeting*.
- Khoo, C. S. G., Chan, S., & Niu, Y. (2000). Extracting causal knowledge from a medical database using graphical patterns. *ACL'00 Proceedings of the 38th Annual Meeting on Association for Computational Linguistics* (pp. 336–343). Morristown, NJ, USA: Association for Computational Linguistics.
- Leroy, G., & Chen, H. (2005). Genescene: An ontology-enhanced integration of linguistic and co-occurrence based relations in biomedical texts. *JASIST*, 56, 457–468.
- Nédellec, C. (2005). Learning language in logic - genic interaction extraction challenge. *In Proceedings of the Learning Language in Logic workshop*.
- Pustejovsky, J., Castano, J., Zhang, J., Cochran, B., & Kotecki, M. (2002). Robust relational parsing over biomedical literature: Extracting inhibit relations. *Pacific Symposium on Biocomputing*.
- Rastier, F., Cavazza, M., & Abeille, A. (2001). *Semantics for descriptions*. Stanford, USA: Center for the Study of Language and Information.
- Ray, S., & Craven, M. (2001). Representing sentence structure in hidden Markov models for information extraction. *In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence* (pp. 1273–1279). Seattle, WA: Morgan Kaufmann.
- Rinaldi, F., Schneider, G., & Kaljurand, K. (2004). Mining relations in the genia corpus. *In "Second European Workshop on Data Mining and Text Mining for Bioinformatics", in conjunction with ECML/PKDD 2004*. Pisa, Italy.
- Sang, E., & Meulder, F. D. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. *In Proceedings of CoNLL'2003*.
- Sekimizu, T., Park, H., & Tsujii, J. (1998). Identifying the interaction between genes and gene products based on frequently seen verbs in medline abstracts. *Genome Informatics*.
- Smith, L. H., Tanabe, L., Rindlesch, T., & Wilbur, W. (2005). Medtag: A collection of biomedical annotations. *In Proceedings of the Joint ACL Workshop and BioLINK SIG (ISMB) on Linking Biological Literature Ontologies and Databases*.
- Stephens, M., Palakal, M., Mukhopadhyay, S., Rajee, R., & Mostafa, J. (2001). Detecting gene relations from medline abstracts. *In Proceedings of the Sixth Annual Pacific Symposium on Biocomputing (PSB 001)*.
- Tanabe, L., Xie, N., L. H. Thom, W. M., & Wilbur, W. J. (2005). Genetag: a tagged corpus for gene/protein named entity recognition. *BMC Bioinformatics*, 6(Suppl 1):S3.
- Thomas, J., Milward, D., Ouzounis, C., & Pulman, S. (2000). Automatic extraction of protein interactions from scientific abstracts. *In Proceedings of Pacific Symposium on Biocomputing*.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann. 2nd edition.
- Zelenko, D., Aone, C., & Richardella, A. (2003). Kernel methods for relation extraction. *J. Mach. Learn. Res.*, 3, 1083–1106.