Exposé for Diploma Thesis

Joint Extraction of Proteins and Bio-Molecular Events using Imperatively Defined Factor Graphs

Tim Rocktäschel
<trocktae@informatik.hu-berlin.de>

Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftliche Fakultät II
Computer Science Department
Knowledge Management in Bioinformatics

Academic advisors:
Prof. Dr. Ulf Leser <leser@informatik.hu-berlin.de>
Dr. Roman Klinger <roman.klinger@scai.fraunhofer.de>
M.Sc. Philippe Thomas <thomas@informatik.hu-berlin.de>

Berlin, November 9, 2011

1 Motivation and Background

Most of the information Protein-Protein Interactions (PPIs) and bio-molecular events can be found in unstructured natural language texts. PubMed, a publicly accessible database for biomedical literature, nowadays consists of over 21 million citations[1] and is still growing fast. Thus, methods that extract information about biological processes automatically from natural language texts are of great importance. They can, for example, lower manual curation efforts of metabolic databases drastically (Ono *et al.*, 2001). For instance, PPIs extracted from natural language texts can be used to automatically build PPI databases or to support curators in building them.

Competitions are held to evaluate the progress achieved by the research community in the extraction of PPIs and bio-molecular events. For instance, the BioCreative II Challenge (Krallinger *et al.*, 2008) focused on the extraction of PPIs. In contrast, the BioNLP Shared Task 2009 (Kim *et al.*, 2009) and 2011 (Kim *et al.*, 2011) both aimed at extracting fine-grained information about detailed behavior of bio-molecules, known as bio-molecular events.

Both, the BioNLP 2009 and 2011 event extraction tasks aimed at extracting the following event types selected from the GENIA ontology (Kim *et al.*, 2003): *gene expression, transcription, protein catabolism, phosphorylation, localization, binding, regulation, positive regulation and negative regulation*. All these events have a protein as their theme (primary argument). Binding has several proteins as its themes. For regulation events, extracting their cause is required. All events have event triggers that need to be extracted as well. Phosphorylation, localization, binding and regulation take a secondary argument (a location or a site). Notably, a regulation event may also take another event as its argument (called nested events).

To separate concerns, in competitions evaluating PPI and bio-molecular event extraction, it is common to use gold standard annotations for all protein mentions. Hence, protein named entity recognition (NER) is not considered and solely the performance for relation extraction (RE) is measured. Kim *et al.* (2009) noted that the only feature which detracts from the BioNLP task's realism is the fact that RE systems were provided with gold standard protein annotations.

Kabiljo *et al.* (2009) showed that the use of gold standard annotations has a high impact on the performance of methods for PPI extraction. This is due to the fact that error propagation caused by NER is neglected. They evaluated AkanePPI (Sætre *et al.*, 2007), a state-of-the-art system for PPI extraction, on five PPI corpora. Kabiljo *et al.* observed a drop in F_1 measure between 2.2 and 22.7 percentage points when protein annotations were discarded and had to be previously identified by a NER system (we refer to this setting as *pipeline architecture*). Thus, the quality of NER has a strong impact on the quality of PPI extraction. The NER system used by Kabiljo *et al.* is BANNER (Leaman and Gonzalez, 2008), which is based on a second order linear-chain Conditional Random Field (CRF) (see Section 4.2) and achieves competitive results in protein name recognition tasks[2].

[1] http://www.ncbi.nlm.nih.gov/pubmed last accessed: November 9, 2011
[2] http://banner.sourceforge.net/ last accessed: November 9, 2011

Clearly, a similar drop in performance can be expected for bio-molecular event extraction when depending on protein NER. One idea to reduce the impact of NER on RE is to perform both steps jointly. We believe that by performing protein NER and bio-molecular event extraction simultaneously, both steps will be able to beneficially influence each other. We expect a performance gain for the joint approach compared to the often used pipeline architecture. This would have a strong impact on the real-life application of bio-molecular event extractors, which rely on NER of proteins and suffer from error propagation.

In this thesis, we want to show that NER and RE can beneficially influence each other. We will employ a probabilistic graphical model for joint extraction, since they can be used to model and query complex joint probabilities. Although we will implement this approach for the example of protein NER and bio-molecular event extraction, the insights gained in this thesis will similarly apply to joint protein NER and PPI extraction. We hope the joint model reduces bio-molecular event extraction errors caused by NER. In addition, event extraction might have a positive influence on the performance of protein NER. Consequently, we will evaluate the joint NER and event extractor in a realistic scenario where no gold standard entities are given and compare it with the pipeline architecture. Furthermore, in contrast to previous work, we aim at extracting all multi-token proteins rather than only single-token proteins that participate in a relation.

1.1 Probabilistic Graphical Models and Factor Graphs

In the following, we will give a brief introduction to probabilistic graphical models based on Koller and Friedman (2009); Bishop (2006); Klinger and Tomanek (2007). Probabilistic graphical models are used to specify joint distributions of random variables. Models are represented as a graph, where the nodes correspond to the variables of the distribution and edges between nodes express probability relationships between these variables (Bishop, 2006, p. 360). More specifically, two random variables are conditionally independent if they are not connected with an edge in the graph (Klinger and Tomanek, 2007). Two random variables $\mathbf{X}$ and $\mathbf{Y}$ are called conditionally independent given $\mathbf{Z}$, if $P(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})P(\mathbf{Y}|\mathbf{Z})$.

One can distinguish between directed graphical models, known as Bayesian Networks (BNs), and undirected graphical models, called Markov Networks (MNs) or Markov Random Fields (MRFs). Only the absence of edges, i.e., the assertion of conditional independence, is informative in graphical models (Klinger and Tomanek, 2007). Hence, they can be seen as a set of independence assumptions. This set of independencies is equivalent with the factorization of the joint distribution into a product of factors (Koller and Friedman, 2009, p. 5).

The factorization of a probabilistic graphical model can be represented by a factor graph, i.e., a bipartite graph where the random variables are connected with their corresponding factors (Kschischang *et al.*, 2001). We shall use FACTORIE (see Section 3.2) to imperatively build undirected graphical models as factor graphs.

In the case of MRFs, a factor Φ is defined as a function from the values of a set of random variables to $\mathbb{R}$. Thus, factors are not probability functions, but measure the compatibility between the variables connected them. However, we get a probability distribution by normalizing the product of factors (Koller and Friedman, 2009, p. 108). Let $\mathbf{X} = X_1, \ldots, X_n$ be the set of random variables, let k be the number of factors that factorizes the distribution and let $\mathbf{D_i}$ be the subset of $\mathbf{X}$ over which factor Φ_i is defined. A MRF factorizes over the distribution with a set $\Phi = \{\Phi_1, \ldots, \Phi_k\}$ of factors, if every $\mathbf{D_i}$ is a clique in the MRF. In other words, all random variables of a factor are fully connected to each other in the MRF. The probability distribution can be obtained by calculating

$$P_{\Phi}(\mathbf{X}) = \frac{\prod_{i=1}^{k} \Phi_i(\mathbf{D_i})}{\sum_{X_1,\ldots,X_n} \prod_{i=1}^{k} \Phi_i(\mathbf{D_i})}.$$

The denominator is a function of $\mathbf{X}$, called normalization constant. As we sum over all possible assignments to the random variables $X_1, \ldots, X_n$, calculating this constant is extremely expensive. However, sampling methods for approximate learning and inference exist, where calculating this constant is not needed (see Metropolis-Hasting Section 4.1.1).

2 Goal

The primary goal of this thesis is to employ a MRF, represented as factor graph, to jointly perform protein NER and bio-molecular event extraction. For this purpose, we will extend the event extraction system introduced by Klinger *et al.* (2011) (see Section 3.2) with protein NER capabilities and train and evaluate it on the BioNLP 2009 corpus.

Secondarily, we will investigate whether a combination of the pipeline architecture and the joint model is beneficial. Therefore, in addition to the protein mentions in the bio-event corpus, we will also incorporate protein annotations from the BioCreative II corpus into the NER component of the joint model by using a custom objective function. Hence, we will address the problem that a bio-event corpus might not be sufficient for training a protein NER component.

3 Related Work

In previous work (Rocktäschel, 2011), we attempted to perform protein NER and PPI extraction jointly by using Support Vector Machines (SVMs) (Joachims, 1998). Thereby we encountered the problem of training SVMs on extremely imbalanced datasets. This resulted in an overall performance decrease compared to the pipeline architecture. With SVMs, we found it labour-intensive and time-consuming to model complex dependencies between different components. In addition, modelling these dependencies with a SVM was costly as it resulted in classifications that were quadratic in the number of tokens for each sentence. Moreover, our approach was limited to single-token proteins. Thus, we could not compare this approach with a pipeline architecture consisting of a state-of-the-art NER and PPI component. Another drawback was the fact that we could solely extract entities that also participated in a PPI. However, joint inference has been successfully applied in several information extraction tasks from different domains.

Kate and Mooney (2010) jointly extracted entities (locations, persons and organizations) and relations (e.g. "works for", "is located in") using a "card-pyramid" graph structure that captures possible entities and relations in a sentence. Their system outperformed a pipeline approach for four out of the five relation types.

The system developed by Zhu *et al.* (2006) extracts records from web pages and jointly labels attributes within a record. It is based on a CRF with a hierarchical graph structure. Their experiments show an improvement for both tasks when performed jointly.

A joint model for parsing and NER was introduced by Finkel and Manning (2009). By using a feature-based CRF-CFG parser and incorporating NER information, they improved the performance of both steps (1.36 percentage points F_1 measure for parsing and 9.0 for NER).

3.1 Markov Logic Networks

Poon and Domingos (2007) extracted citations from unstructured natural language texts by jointly segmenting records and matching them with those records that refer to the same entity (entity resolution). They used a Markov Logic Network (MLN) (Richardson and Domingos, 2006), which is a set of weighted first order logic formulas with which one can define predicates to describe relations. The formulas representing the MLN are then used to instantiate a MRF. By employing a MLN, their system was able to outperform all previous approaches.

In the BioNLP 2009 core event extraction task, Riedel *et al.* (2009) obtained the 4th best F_1 measure of 44.4% with a MLN. The MLN was used to jointly predict triggers and arguments of bio-molecular events, whereas a common approach is to use several classifiers in a pipeline architecture. Riedel *et al.* used predicates to link arguments and triggers in the dependency tree. Furthermore, they used formulas to model constraints that every extracted event should satisfy. Later, a similar system based on a MLN was developed by Poon and Vanderwende (2010) for this task, achieving an F_1 measure of 50.0%. Note that these approaches focused on joint extraction of triggers and events given protein gold standard annotations, whereas we also aim at jointly extracting protein mentions.

A modification (Riedel and McCallum, 2011) of this system achieved the 2nd best F_1 measure (55.2%) in the BioNLP 2011 Genia Event Task. By combining this system with event parsing, the resulting stacked system (Riedel *et al.*, 2011) yielded the best F_1 measure (56.0%) among all participants in this task.

3.2 Imperatively Defined Factor Graphs

McCallum *et al.* (2009, 2008) introduced FACTORIE, a library for implementing Imperatively Defined Factor Graphs (IDFs). Arbitrary factor graphs (e.g. for MRFs) can be imperatively constructed with FACTORIE by defining variables and factor templates, i.e., factors that are tied to the same parameters. The resulting factor graph is called a IDF. Furthermore, so-called user-defined proposal (see Section 4.1.1) and objective functions (see Section 4.1.2) can be implemented to improve the resulting model and to speed up training and inference. In experiments on joint segmentation and co-reference of research paper citations, McCallum *et al.* achieved an error reduction by $20 - 25\%$ and were 3 to 15 times faster than a MLN. Besides this performance advantage, IDFs allow more accurate and expressive modeling than MLNs (McCallum *et al.*, 2009).

FACTORIE has already been employed successfully for various other tasks: For instance, Wick *et al.* (2010) implemented a probabilistic database with it. A phrase-based machine translation system modeled as CRF was introduced by Roth *et al.* (2010). Yao *et al.* (2010) used FACTORIE for relation extraction. They applied joint inference and distant supervision, i.e., they trained their model on a knowledge base rather than a gold standard.

Klinger *et al.* (2011) developed a cross-sentence event extractor modeled as factor graph and reported competitive results on the BioNLP 2009 Shared Task. Their key idea was to incorporate information about event-event dependencies. These dependencies were used to model the observation that events are introduced across sentences in a discourse of a paper. Thus, extracting a particular event should have an influence on extracting following events in that paper.

4 Approach

As mentioned in Section 2, we will build a joint factor graph by extending the event extraction system developed by Klinger *et al.* (2011) with protein NER capabilities. The resulting model will be used to jointly extract proteins and bio-molecular events from natural language texts.

4.1 Markov Chain Monte Carlo Methods

Factor graphs can become extremely large due to complex variable dependencies, rich feature sets and large training corpora. Inference and parameter estimation (learning) becomes intractable in large factor graphs. Fortunately, Markov Chain Monte Carlo (MCMC) (Hastings, 1970) can be used to draw samples from the high-dimensional probability distribution p encoded in a factor graph. Subsequently, these samples can be used to calculate marginal and conditional probabilities and, thus, to answer various kinds of queries (Richardson and Domingos, 2006). To obtain accurate models with FACTORIE, the principle of MCMC-sampling using a proposal function and learning with an objective function is essential. The former can be achieved with Metropolis-Hastings and the latter with SampleRank.

A Markov Chain is defined over a state space where each state corresponds to an assignment of values to the variables in the graphical model. Additionally, a transition model is used to specify the probability of going from one state to another (i.e. assigning other values to the variables). The basic idea of MCMC is to perform random walks over Markov Chains to sample from the high-dimensional probability distribution p underlying a graphical model (Koller and Friedman, 2009, p. 507). If the Markov Chain fulfils certain properties (if it is irreducible and aperiodic), it will converge to a unique stationary distribution π (Brooks, 1998). Thus, after we have constructed a Markov Chain of a certain length, all samples generated from this Markov Chain will be sampled from its stationary distribution.

The key is to ensure that the stationary distribution π is the desired distribution p (Gilks and Richardson, 1996, p. 5). This can be achieved by a suitable transition model for the Markov Chain.

4.1.1 Metropolis-Hastings

A very popular MCMC method that ensures that the Markov Chain converges to a desired stationary distribution p is Metropolis-Hastings (MH) (Hastings, 1970). The following explanations are based on (Koller and Friedman, 2009, pp. 516-518) and (Andrieu *et al.*, 2003, pp. 13-16).

The main advantage of MH is the fact that we only have to know p up to a normalization constant. We do not have to sample from p directly, but instead use another distribution q, called proposal distribution, to generate samples. In fact, we do not use q to generate actual samples, but instead we only propose transitions for the Markov Chain based on q. These proposals can be seen as a transition model T^Q, where $T^Q(x, x')$ denotes the probability proposing a change from x to x' (we shall call T^Q a *proposer* or *sampler*). Subsequently, MH will correct the error made by not sampling from p. Thus, in the end the Markov Chain will generate samples that are distributed as if we would have drawn them directly from p. However, MH does not keep track of the error corrections while constructing the Markov Chain. Instead, MH randomly decides to accept or reject a proposal made by T^Q based on an acceptance probability A. This probability specifies for every pair (x, x') whether to accept the proposed transition to x' or to stay at x. By defining a transition model T and an acceptance probability A in the following way, the resulting Markov Chain will have p as its stationary distribution:

$$A(x, x') = \min\left[1, \frac{p(x')T^Q(x', x)}{p(x)T^Q(x, x')}\right]$$

$$T(x, x') = T^Q(x, x')A(x, x') \quad \text{for } x \neq x'$$

$$T(x, x) = T^Q(x, x) \sum_{x \neq x'} T^Q(x, x')(1 - A(x, x'))$$

As we choose a T^Q that is easy to calculate, the ratio $T^Q(x', x)/T^Q(x, x')$ will also be easy to calculate. In the context of MRFs, both $p(x')$ and $p(x)$ can be calculated using only the factors that are connected to the variable which T^Q proposed to change. Note that because p is used in the numerator and denominator, it is sufficient to know p up to a normalization constant. To obtain the ratio $p(x')/p(x)$, we can thus simply multiply the corresponding factors without normalizing them to actual probabilities.

4.1.2 FACTORIE's SampleRank

SampleRank (Wick *et al.*, 2009, 2011) is a MCMC-based parameter estimation algorithm used in FACTORIE to train the parameters of factor graphs. Parameter updates are performed after each MCMC step. While exploring the parameter space, SampleRank ranks local moves in this space using an objective function. A main feature of SampleRank is the fact that we can incorporate user-defined objective functions, which results in faster training and more accurate models. For instance, we can use an objective function correlated to F_1, which probably results in a model that will be more balanced between precision and recall on the evaluation data then a model using, for instance, per-token accuracy.

4.2 Named Entity Recognition of Proteins

We will use FACTORIE to model a factor graph for protein NER. The underlying graphical model will be a Conditional Random Field (CRF). A CRF is a MRF which directly defines the conditional probability $p(\mathbf{Y}|\mathbf{X})$, where $\mathbf{X}$ is a set of input variables (observation) and $\mathbf{Y}$ is a set of output variables (Lafferty *et al.*, 2001). CRFs have a great advantage over generative models (e.g. Hidden Markov Models (HMMs)) as they do not have to make assumptions about the underlying observation distribution $\mathbf{X}$. Furthermore, a huge number of non-independent features can be used to describe the input data (McCallum *et al.*, 2000). CRFs have become the de facto standard for machine-learning-based NER methods over the last decade.

To achieve near state-of-the-art performance, we will use common morphological, orthographical and syntax features for protein NER provided by BANNER (Leaman and Gonzalez, 2008). When using FACTORIE, we have to implement a proposer (see Section 4.1.1) for protein mentions in text. Furthermore, we will experiment with user-defined objective functions (see Section 4.1.2). Subsequently, we can compare the performance of our system with BANNER by training it on the BioCreative II Gene Mention Tagging[3] training set and evaluating it on the test set.

Additionally, we will measure the impact of protein NER on the performance of Klinger *et al.* (2011)'s bio-event extraction system using a pipeline architecture. Hence, the bio-event extraction system will only rely on the proteins sampled by our protein NER proposer instead of gold standard proteins. The pipeline will be trained on the training corpus of the BioNLP 2009 Shared Task and evaluated on the development corpus. We expect a performance drop for bio-event extraction similar to the severe drop observed by Kabiljo *et al.* (2009) for PPI extraction when relying on protein NER rather than gold standard entities.

4.3 Joint Extraction of Proteins and Bio-Molecular Events

The primary goal of this thesis is the joint extraction of proteins and bio-molecular events. Our proposer for protein NER and the proposers included in Klinger *et al.*'s bio-event extractor will be used to simultaneously propose proteins and bio-molecular events. By applying a Metropolis-Hastings-like MCMC algorithm (see Section 4.1.1), FACTORIE jointly accepts or rejects this proposal, resulting in joint learning and joint inference.

We will compare the joint extractor with the pipeline architecture. Here, we expect a performance gain for both, the bio-event extraction and protein NER, since they will probably be able to beneficially influence each other when performed jointly.

[3] http://biocreative.sourceforge.net/biocreative_2_gm.html last accessed: November 9, 2011

4.4 Training on two corpora

Since the joint extractor is trained on an event-extraction corpus, the protein proposer will, during training, only have access to proteins contained in this corpus. This might result in a low-performing model.

SampleRank only evaluates variables whose value were changed by the proposer. Thus, different objective functions can be used for different stages in the process of joint extraction. For instance, by specifying an objective function that has access to other corpora (e.g. the BioCreative II corpus), it should be possible to provide the protein proposer with additional training data.

The resulting model will still be able to perform joint extraction of proteins and bio-events. However, additional training data should improve the protein NER. Hence, this could have a positive effect on both, the quality protein NER and, subsequently, also the quality of bio-molecular event extraction. Implementing and evaluating such a model is a secondary goal of this thesis. However, its feasibility and performance-advantages are rather unclear.

References

Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003). An Introduction to MCMC for Machine Learning. *Machine Learning*, **50**(1), 5–43.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, volume 4 of *Information science and statistics*. Springer.

Brooks, S. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **47**(1), 69–100.

Finkel, J. R. and Manning, C. D. (2009). Joint parsing and named entity recognition. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics on - NAACL '09*.

Gilks, W. and Richardson, S. (1996). *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine Learning ECML98*, **1398**(23), 137–142.

Kabiljo, R., Clegg, A. B., and Shepherd, A. J. (2009). A realistic assessment of methods for extracting gene/protein interactions from free text. *BMC Bioinformatics*, **10**(1).

Kate, R. and Mooney, R. (2010). Joint entity and relation extraction using card-pyramid parsing. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 203–212. Association for Computational Linguistics.

Kim, J., Ohta, T., Pyysalo, S., and Kano, Y. (2009). Overview of BioNLP'09 shared task on event extraction. *Proceedings of the Workshop on BioNLP: Shared Task*, (June), 1–9.

Kim, J., Pyysalo, S., Ohta, T., Bossy, R., and Nguyen, N. (2011). Overview of BioNLP Shared Task 2011. *ACL HLT 2011*, **2009**.

Kim, J.-D., Ohta, T., Tateisi, Y., and Tsujii, J. (2003). GENIA corpus–a semantically annotated corpus for bio-textmining. *Bioinformatics*, **19**(Suppl 1), i180–i182.

Klinger, R. and Tomanek, K. (2007). Classical Probabilistic Models and Conditional Random Fields. Technical Report December, Technische Universität Dortmund.

Klinger, R., Riedel, S., and McCallum, A. (2011). Inter-Event Dependencies support Event Extraction from Biomedical Literature. In *Mining Complex Entities from Network and Biomedical Data (MIND), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models*. MIT Press.

Krallinger, M., Leitner, F., Rodriguez-Penagos, C., and Valencia, A. (2008). Overview of the protein-protein interaction annotation extraction task of BioCreative II. *Genome Biology*, **9**(Suppl 2).

Kschischang, F., Frey, B., and Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, **47**(2), 498–519.

Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data. *ICML-2001*, **pages**, 282–289.

Leaman, R. and Gonzalez, G. (2008). BANNER: an executable survey of advances in biomedical named entity recognition. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*.

McCallum, A., Freitag, D., and Pereira, F. (2000). Maximum entropy Markov models for information extraction and segmentation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 591–598.

McCallum, A., Rohanemanesh, K., Wick, M., Schultz, K., and Singh, S. (2008). FACTORIE: Efficient probabilistic programming for relational factor graphs via imperative declarations of structure, inference and learning. In *Neural Information Processing Systems (NIPS), Workshop on Probabilistic Programming*, pages 1–3.

McCallum, A., Schultz, K., and Singh, S. (2009). FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs. *Neural Information Processing Systems (NIPS)*, **22**, 1–9.

Ono, T., Hishigaki, H., Tanigami, A., and Takagi, T. (2001). Automated extraction of information on protein-protein interactions from the biological literature. *Bioinformatics*.

Poon, H. and Domingos, P. (2007). Joint Inference in Information Extraction. In *Proceedings Of The National Conference On Artificial Intelligence*, volume 22.

Poon, H. and Vanderwende, L. (2010). Joint Inference for Knowledge Extraction from Biomedical Literature. In *Computational Linguistics*, number June, pages 813–821. Association for Computational Linguistics.

Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine Learning*, **62**(1-2), 107–136.

Riedel, S. and McCallum, A. (2011). Robust Biomedical Event Extraction with Dual Decomposition and Minimal Domain Adaptation. *Proceedings of the Natural Language Processing in Biomedicine NAACL 2011 Workshop (BioNLP '11)*, pages 46–50.

Riedel, S., Chun, H., and Takagi, T. (2009). A markov logic approach to bio-molecular event extraction. *Proceedings of the Workshop on BioNLP: Shared Task*, pages 41–49.

Riedel, S., McClosky, D., Surdeanu, M., McCallum, A., and Manning, C. D. (2011). Model combination for event extraction in bionlp 2011. *Proceedings of the Natural Language Processing in Biomedicine NAACL 2011 Workshop (BioNLP '11*, pages 51–55.

Rocktäschel, T. (2011). Single-Step Extraction of Protein-Protein Interactions with Support Vector Machines. *Studienarbeit. Humboldt-Universität zu Berlin*.

Roth, B., McCallum, A., Dymetman, M., and Cancedda, N. (2010). Machine Translation Using Overlapping Alignments and SampleRank. *Proceedings of Association for Machine Translation in the Americas, Denver, CO*.

Sætre, R., Yoshida, K., Yakushiji, A., Miyao, Y., Matsubyashi, Y., and Ohta, T. (2007). AKANE system: protein-protein interaction pairs in BioCreAtIvE2 challenge, PPI-IPS subtask. In *Proceedings of the Second BioCreative Challenge Workshop*, pages 209–212. Citeseer.

Wick, M., Rohanimanesh, K., Culotta, A., and McCallum, A. (2009). SampleRank: Learning Preferences from Atomic Gradients. In *Neural Information Processing Systems (NIPS), Workshop on Advances in Ranking*, pages 1–5.

Wick, M., McCallum, A., and Miklau, G. (2010). Scalable Probabilistic Databases with Factor Graphs and MCMC. *Proceedings of the VLDB Endowment*, **3**(1-2), 794–804.

Wick, M., Rohanimanesh, K., Bellare, K., Culotta, A., and Mccallum, A. (2011). SampleRank: Training Factor Graphs with Atomic Gradients. *Proceedings of the International Conference on Machine Learning (ICML)*.

Yao, L., Riedel, S., and McCallum, A. (2010). Collective Cross-Document Relation Extraction Without Labelled Data. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, (October), 1013–1023.

Zhu, J., Nie, Z., Wen, J.-R., Zhang, B., and Ma, W.-Y. (2006). Simultaneous record detection and attribute labeling in web data extraction. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*.