Semantic Consistency: A Local Subspace Based Method for Distant Supervised Relation Extraction

Xianpei Han and **Le Sun**
State Key Laboratory of Computer Science
Institute of Software, Chinese Academy of Sciences
HaiDian District, Beijing, China.
{xianpei, sunle}@nfs.iscas.ac.cn

Abstract

One fundamental problem of distant supervision is the noisy training corpus problem. In this paper, we propose a new distant supervision method, called *Semantic Consistency*, which can identify reliable instances from noisy instances by inspecting whether an instance is located in a semantically consistent region. Specifically, we propose a *semantic consistency* model, which first models the local subspace around an instance as a sparse linear combination of training instances, then estimate the semantic consistency by exploiting the characteristics of the local subspace. Experimental results verified the effectiveness of our method.

1 Introduction

Relation extraction aims to identify and categorize relations between pairs of entities in text. Due to the time-consuming annotation process, one critical challenge of relation extraction is the lack of training data. To address this limitation, a promising approach is *distant supervision (DS)*, which can automatically gather labeled data by heuristically aligning entities in text with those in a knowledge base (Mintz et al., 2009). The underlying assumption of distant supervision is that every sentence that mentions two entities is likely to express their relation in a knowledge base.

Relation Instance	Label
S1: *Jobs* *was the founder of* ***Apple***	Founder-of, CEO-of
S2: *Jobs* *joins* ***Apple***	Founder-of, CEO-of

Figure 1. Labeled instances by distant supervision, using relations *CEO-of(Steve Jobs, Apple Inc.)* and *Founder-of(Steve Jobs, Apple Inc.)*

The distant supervision assumption, unfortunately, can often fail and result in a noisy training corpus. For example, in Figure 1 DS assumption will wrongly label S1 as a *CEO-of* instance and S2 as instance of *Founder-of* and *CEO-of*. The noisy training corpus in turn will lead to noisy extractions that hurt extraction accuracy (Riedel et al., 2010).

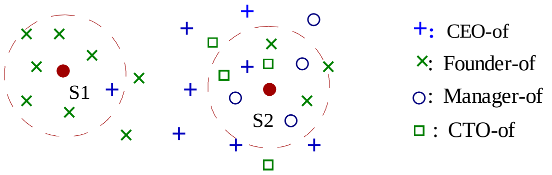

Figure 2. The regions the two instances in Figure 1 located, where: 1) S1 locates in a semantically consistent region; and 2) S2 locates in a semantically inconsistent region

To resolve the noisy training corpus problem, this paper proposes a new distant supervision method, called *Semantic Consistency*, which can effectively identify reliable instances from noisy instances by inspecting whether an instance is located in a semantically consistent region. Figure 2 shows two intuitive examples. We can see that, semantic consistency is an effective way to identify reliable instances. For example, in Figure 2 S1 is highly likely a reliable *Founder-of* instance because its neighbors are highly semantically consistent, i.e., most of them express the same relation type – *Founder-of*. On contrast S2 is highly likely a noisy instance because its neighbors are semantically inconsistent, i.e., they have a diverse relation types. The problem now is how to model the semantic consistency around an instance.

To model the semantic consistency, this paper proposes a local subspace based method. Specifically, given sufficient training instances, our method first models each relation type as a linear subspace spanned by its training instances. Then, the local subspace around an instance is modeled and characterized by seeking the sparsest linear combination of training instances which can reconstruct the instance. Finally, we estimate the semantic consistency of an instance by exploiting the characteristics of its local subspace.

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 718–724,
Baltimore, Maryland, USA, June 23-25 2014. ©2014 Association for Computational Linguistics

This paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed method. Section 4 presents the experiments. Finally Section 5 concludes this paper.

2 Related Work

This section briefly reviews the related work. Craven and Kumlien (1999), Wu et al. (2007) and Mintz et al.(2009) were several pioneer work of distant supervision. One main problem of DS assumption is that it often will lead to false positives in training data. To resolve this problem, Bunescu and Mooney (2007), Riedel et al. (2010) and Yao et al. (2010) relaxed the DS assumption to the *at-least-one* assumption and employed multi-instance learning techniques to identify wrongly labeled instances. Takamatsu et al. (2012) proposed a generative model to eliminate noisy instances.

Another research issue of distant supervision is that a pair of entities may participate in more than one relation. To resolve this problem, Hoffmann et al. (2010) proposed a method which can combine a sentence-level model with a corpus-level model to resolve the multi-label problem. Surdeanu et al. (2012) proposed a multi-instance multi-label learning approach which can jointly model all instances of an entity pair and all their labels. Several other research issues also have been addressed. Xu et al. (2013), Min et al. (2013) and Zhang et al. (2013) try to resolve the false negative problem raised by the incomplete knowledge base problem. Hoffmann et al. (2010) and Zhang et al. (2010) try to improve the extraction precision by learning a dynamic lexicon.

3 The Semantic Consistency Model for Relation Extraction

In this section, we describe our semantic consistency model for relation extraction. We first model the subspaces of all relation types in the original feature space, then model and characterize the local subspace around an instance, finally estimate the semantic consistency of an instance and exploit it for relation extraction.

3.1 Testing Instance as a Linear Combination of Training Instances

In this paper, we assume that there exist *k* distinct relation types of interest and each relation type is represented with an integer index from *1* to *k*. For *i*th relation type, we assume that totally n_i training instances $V_i = \{\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, ..., \mathbf{v}_{i,n_i}\}$ have been collected using DS assumption. And each instance is represented as a weighted feature vector, such as the features used in (Mintz, 2009) or (Surdeanu et al., 2012), with each feature is TFIDF weighted by taking each instance as an individual document.

To model the subspace of *i*th relation type in the original feature space, a variety of models have been proposed to discover the underlying patterns of V_i. In this paper, we make a simple and effective assumption that *the instances of a single relation type can be represented as the linear combination of other instances of the same relation type*. This assumption is well motived in relation extraction, because although there is nearly unlimited ways to express a specific relation, in many cases basic principles of economy of expression and/or conventions of genre will ensure that certain systematic ways will be used to express a specific relation (Wang et al., 2012). For example, as shown in (Hearst, 1992), the *IS-A* relation is usually expressed using several regular patterns, such as "*such NP as {NP ,}* {(or | and)} NP*" and "*NP {, NP}* {,} or other NP*".

Based on the above assumption, we hold many instances for each relation type and directly use these instances to model the subspace of a relation type. Specifically, we represent an instance ***y*** of *i*th type as the linear combination of training instances associated with *i*th type:

$$\mathbf{y} = \alpha_{i,1}\mathbf{v}_{i,1} + \alpha_{i,2}\mathbf{v}_{i,2} + ... + +\alpha_{i,n_i}\mathbf{v}_{i,n_i} \quad (1)$$

for some scalars $\alpha_{i,j}$, with *j = 1, 2, ...,*n_i. For example, we can represent the *CEO-of* instance "***Jobs** was the CEO of **Apple***" as the following linear combination of *CEO-of* instances:

- 0.8: ***Steve Ballmer** is the CEO of **Microsoft***
- 0.2: ***Rometty** was served as the CEO of **IBM***

For simplicity, we arrange the given n_i training instances of *i*th relation type as columns of a matrix $A_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, ..., \mathbf{v}_{i,n_i}]$, then we can write the matrix form of Formula 1 as:

$$\mathbf{y} = \mathbf{A}_i\mathbf{x}_i \quad (2)$$

where $\mathbf{x}_i = [\alpha_{i,1},\ ...,\ \alpha_{i,n_i}]$ is the coefficient vector. In this way, the subspace of a relation type is the linear subspace spanned by its training instances, and if we can find a valid $\mathbf{x_i}$, we can explain ***y*** as a valid instance of *i*th relation type.

3.2 Local Subspace Modeling via Sparse Representation

Based on the above model, the local subspace of an instance is modeled as the linear combination of training instances which can reconstruct the instance. Specifically, to model the local subspace, we first concatenate the *n* training instances of all *k* relation types:

$$A = [A_1, A_2, ..., A_k]$$

Then the local subspace around ***y*** is modeled by seeking the solution of the following formula:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad (3)$$

However, because of the redundancy of training instances, Formula 3 usually has more than one solution. In this paper, following the idea in (Wright et al., 2009) for robust face recognition, we use the sparsest solution (i.e., how to reconstruct an instance using minimal training instances), which have been shown is both discriminant and robust to noisiness. Concretely, we seek the sparse linear combination of training instances to reconstruct ***y*** by solving:

$$(l^1) : \mathbf{x}^* = \arg\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \varepsilon \quad (4)$$

where $\mathbf{x} = [\alpha_{1,1}, ..., \alpha_{1,n_1}, ..., \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}, ...]$ is a coefficient vector which identifies the spanning instances of ***y***'s local subspace, i.e., the instances whose $\alpha_{i,j} \neq 0$. In practice, the training corpus may be too large to direct solve Formula 4. Therefore, this paper uses the K-Nearest-Neighbors (KNN) of **y** (*1000* nearest neighbors in this paper) to construct the training instance matrix **A** for each **y**, and KNN can be searched very efficiently using specialized algorithms such as the LSH functions in (Andoni & Indyk, 2006).

Through the above semantic decomposition, we can see that, the entries of **x** can encode the underlying semantic information of instance ***y***. For *i*th relation type, let $\delta_i(\mathbf{x})$ be a new vector whose only nonzero entries are the entries in ***x*** that are associated with *i*th relation type, then we can compute the semantic component corresponding to *i*th relation type as $\mathbf{y}_i = \mathbf{A}\delta_i(\mathbf{x})$. In this way a testing instance **y** will be decomposed into *k* semantic components, with each component corresponds to one relation type (with an additional noise component ϵ):

$$\mathbf{y} = \mathbf{y}_1 + ... + \mathbf{y}_i + ... + \mathbf{y}_k + \epsilon \qquad (5)$$

$$\mathrm{S1} = 0.8 \times \underset{\textbf{Founder-of}}{\begin{bmatrix} \text{was} \\ \text{co-founder} \\ \text{of} \\ ... \end{bmatrix}} + 0.2 \times \underset{\textbf{noise}}{\begin{bmatrix} \text{Jobs} \\ \text{Apple} \\ \text{the} \\ ... \end{bmatrix}}$$

$$\mathrm{S2} = 0.1 \underset{\textbf{Founder-of}}{\begin{bmatrix} \text{join} \\ ... \end{bmatrix}} + 0.1 \underset{\textbf{CEO-of}}{\begin{bmatrix} \text{join} \\ ... \end{bmatrix}} + 0.1 \underset{\textbf{CTO-of}}{\begin{bmatrix} \text{join} \\ ... \end{bmatrix}} + ...$$

Figure 3. The semantic decomposition of the two instances in Figure 1

Figure 3 shows an example of semantic decomposition. We can see that, the semantic decomposition can effectively summarize the semantic consistency information of ***y***'s local subspace: if the instances around an instance have diverse relation types (S2 for example), its information will be scattered on many different semantic components. On contrast if the instances around an instance have consistent relation types (S1 for example), most of its information will concentrate on the corresponding relation type.

3.3 Semantic Consistency based Relation Extraction

This section describes how to estimate and exploit the semantic consistency for relation extraction. Specifically, given **y**'s semantic decomposition:

$$\mathbf{y} = \mathbf{y}_1 + ... + \mathbf{y}_i + ... + \mathbf{y}_k + \epsilon$$

we observe that if instance **y** locates at a semantic consistent region, then all its information will concentrate on a specific component $\mathbf{y}_\mathbf{i}$, with all other components equal to zero vector **0**. However, modeling errors, expression ambiguity and noisy features will lead to small nonzero components. Based on the above discussion, we define the semantic consistency of an instance as the semantic concentration degree of its decomposition:

Definition 1(Semantic Consistency). *For an instance* **y***, its semantic consistency with i*th *relation type is:*

$$\text{Consistency}(\mathbf{y}, i) = \frac{\|\mathbf{y}_i\|_2}{\sum_i \|\mathbf{y}_i\|_2 + \|\epsilon\|_2}$$

where Consistency(**y**, i)∈ $[0, 1]$ and will be *1.0* if all information of ***y*** is consistent with *i*th relation type; on contrast it will be *0* if no information in ***y*** is consistent with *i*th relation type.

Semantic Consistency based Relation Extraction. To get accurate extractions, we determine the relation type of ***y*** based on both: 1) How much information in ***y*** is related to *i*th type; and 2) its semantic consistency score with *i*th type, i.e., whether ***y*** is a reliable instance of *i*th type.

To measure how much information in ***y*** is related to *i*th relation type, we compute the proportion of common information between ***y*** and ***y**ᵢ*:

$$\text{sim}(\mathbf{y}, \mathbf{y}_i) = \frac{\mathbf{y} \cdot \mathbf{y}_i}{\mathbf{y} \cdot \mathbf{y}} \qquad (6)$$

Then the likelihood for a testing instance ***y*** expressing *i*th relation type is scored by summarizing both its information and semantic consistency:

$$\text{rel}(\mathbf{y}, i) = \text{sim}(\mathbf{y}, \mathbf{y}_i) \times \text{Consistency}(\mathbf{y}, i)$$

and **y** will be classified into *i*th relation type if its likelihood is larger than a threshold:

$$\text{rel}(\mathbf{y}, i) \ \ge \ \tau_i \qquad (7)$$

where τ_i is a relation type specific threshold learned from training dataset.

Multi-Instance Evidence Combination. It is often that an entity pair will match more than one sentence. To exploit such redundancy for more confident extraction, this paper first combines the evidence from different instances by combing their underlying components. That is, given the matched *m* instances Y={$\mathbf{y}^1$, $\mathbf{y}^2$, **…**, $\mathbf{y}^m$} for an entity pair (e_1, e_2), we first decompose each instance as $\mathbf{y}^j = \mathbf{y}_1^j + ... + \mathbf{y}_k^j + \epsilon$, then the entity-pair level decomposition $\mathbf{y} = \mathbf{y}_1 + ... + \mathbf{y}_k + \epsilon$ is obtained by summarizing semantic components of different instances: $\mathbf{y_i} = \sum_{1 \leq j \leq m} \mathbf{y}_i^j$. Finally, the likelihood of an entity pair expressing *i*th relation type is scored as:

$$\mathrm{rel}(\mathrm{Y}, i) = \mathrm{sim}(\mathbf{y}, \mathbf{y}_i)\mathrm{Consistency}(\mathbf{y}, i)\log(m+1)$$

where $\log(m+1)$ is a score used to encourage extractions with more matching instances.

3.4 One further Issue for Distant Supervision: Training Instance Selection

The above model further provides new insights into one issue for distant supervision: *training instance selection*. In this paper, we select informative training instances by seeking a most compact subset of instances which can span the whole subspace of a relation type. That is, all instances of *i*th type can be represented as a linear combination of these selected instances.

However, finding the optimal subset of training instances is difficult, as there exist 2^N possible solutions for a relation type with *N* training instances. Therefore, this paper proposes an approximate training instance selection algorithm as follows:

1) Computing the centroid of *i*th relation type as $\mathbf{v}_i = \sum_{1 \leq j \leq n_i} \mathbf{v}_{i,j}$
2) Finding the set of training instances which can most compactly span the centroid by solving:
   $(l^1) : \mathbf{x}_i = \arg\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{A}_i\mathbf{x} - \mathbf{v}_i\|_2 \leq \varepsilon$
3) Ranking all training instances according to their absolute coefficient weight value $|\alpha_{i,j}|$;
4) Selecting top *p* percent ranked instances as final training instances.

The above training instance selection has two benefits. First, it will select informative instances and remove redundant instances: an informative instance will receive a high $|\alpha_{i,j}|$ value because many other instances can be represented using it; and if two instances are redundant, the sparse solution will only retain one of them. Second, most of the wrongly labeled training instances will be filtered, because these instances are usually not regular expressions of *i*th type, so they appear only a few times and will receive a small $|\alpha_{i,j}|$.

4 Experiments

In this section, we assess the performance of our method and compare it with other methods.

Dataset. We assess our method using the KBP dataset developed by Surdeanu et al. (2012). The KBP is constructed by aligning the relations from a subset of English Wikipedia infoboxes against a document collection that merges two distinct sources: (1) a 1.5 million documents collection provided by the KBP shared task(Ji et al., 2010; Ji et al., 2011); and (2) a complete snapshot of the June 2010 version of Wikipedia. Totally 183,062 training relations and 3,334 testing relations are collected. For tuning and testing, we used the same partition as Surdeanu et al. (2012): 40 queries for development and 160 queries for formal evaluation. In this paper, each instance in KBP is represented as a feature vector using the features as the same as in (Surdeanu et al., 2012).

Baselines. We compare our method with four baselines as follows:

- ***Mintz++***. This is a traditional DS assumption based model proposed by Mintz et al.(2009).
- ***Hoffmann***. This is an *at-least-one* assumption based multi-instance learning method proposed by Hoffmann et al. (2011).
- ***MIML***. This is a multi-instance multi-label model proposed by Surdeanu et al. (2012).
- ***KNN***. This is a classical *K-Nearest-Neighbor* classifier baseline. Specifically, given an entity pair, we first classify each matching instance using the labels of its 5 (tuned on training corpus) nearest neighbors with cosine similarity, then all matching instances' classification results are added together.

Evaluation. We use the same evaluation settings as Surdeanu et al. (2012). That is, we use the official KBP scorer with two changes: (a) relation mentions are evaluated regardless of their support document; and (b) we score only on the subset of gold relations that have at least one mention in matched sentences. For evaluation, we use *Mintz++*, *Hoffmann*, and *MIML* implementation from Stanford's MIMLRE package (Surdeanu et al., 2012) and implement *KNN* by ourselves.

4.1 Experimental Results

4.1.1 Overall Results

We conduct experiments using all baselines and our semantic consistency based method. For our method, we use top 10% weighted training instances. All features occur less than 5 times are filtered. All l^1-minimization problems in this paper are solved using the augmented Lagrange multiplier algorithm (Yang et al., 2010), which has been proven is accurate, efficient, and robust. To select the classification threshold τ_i for *i*th relation type, we use the value which can achieve the best F-measure on training dataset (with an additional restriction that precision should > 10%).

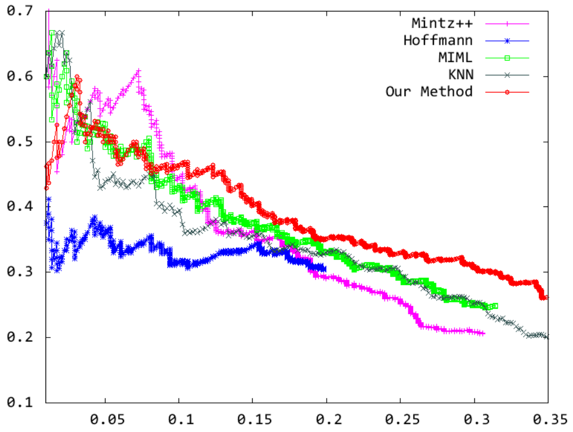

Figure 4. Precision/recall curves in KBP dataset

System	Precision	Recall	F1
Mintz++	0.260	0.250	0.255
Hoffmann	0.306	0.198	0.241
MIML	0.249	0.314	0.278
KNN	0.261	0.295	0.277
Our method	0.286	0.342	0.311

Table 1. The best F1-measures in KBP dataset

Figure 4 shows the precision/recall curves of different systems, and Table 1 shows their best F1-measures. From these results, we can see that:

1) The semantic consistency based method can achieve robust and competitive performance: in KBP dataset, our method correspondingly achieves 5.6%, 7%, 3.3% and 3.4% F1 improvements over the *Mintz++*, *Hoffmann*, *MIML* and *KNN* baselines. We believe this verifies that the semantic consistency around an instance is an effective way to identify reliable instances.

2) From Figure 4 we can see that our method achieves a consistent improvement on the high-recall region of the KBP curves (when recall > *0.1*). We believe this is because by modeling the semantic consistency using the local subspace around each testing instance, our method can better solve the classification of long tail instances which are not expressed using salient patterns.

3) The local subspace around an instance can be effectively modeled as a linear subspace spanned by training instances. From Table 1 we can see that both our method and *KNN* baseline (where the local subspace is spanned using its k nearest neighbors) achieve competitive performance: even the simple *KNN* baseline can achieve a competitive performance (0.277 in F1). This result shows: a) the effectiveness of instance-based subspace modeling; and b) by partitioning subspace into many local subspaces, the subspace model is more adaptive and robust to model prior.

4) The sparse representation is an effective way to model the local subspace using training instances. Compared with *KNN* baseline, our method can achieve a 3.4% F1 improvement. We believe this is because: (1) the discriminative nature of sparse representation as shown in (Wright et al., 2009); and (2) the sparse representation globally seeks the combination of training instances to characterize the local subspace, on contrast *KNN* uses only its nearest neighbor in the training data, which is more easily affected by noisy training instances(e.g., false positives).

4.1.2 Training Instance Selection Results

To demonstrate the effect of training instance selection, Table 2 reports our method's performance using different proportions of training instances.

Proportion	5%	10%	20%	100%
Best F1	0.282	0.311	0.305	0.280

Table 2. The best F1-measures using different proportions of top weighted training instances

From Table 2, we can see that: ① Our training instance selection algorithm is effective: our method can achieve performance improvement using only top weighted instances. ② The training instances are highly redundant: using only 10% weighted instances can achieve a competitive performance.

5 Conclusion and Future Work

This paper proposes a semantic consistency method, which can identify reliable instances from noisy instances for distant supervised relation extraction. For future work, we want to design a more effective instance selection algorithm and embed it into our extraction framework.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grants no. 61100152 and 61272324, and the National High Technology Development 863 Program of China under Grants no. 2013AA01A603.

Reference

Andoni, Alexandr, and Piotr Indyk . 2006. *Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions*. In: Foundations of Computer Science, 2006, pp. 459-468.

Bunescu, Razvan, and Raymond Mooney. 2007. *Learning to extract relations from the web using minimal supervision*. In: ACL 2007, pp. 576.

Craven, Mark, and Johan Kumlien. 1999. *Constructing biological knowledge bases by extracting information from text sources*. In : Proceedings of AAAI 1999.

Downey, Doug, Oren Etzioni, and Stephen Soderland. 2005. *A probabilistic model of redundancy in information extraction*, In: Proceeding of IJCAI 2005.

Gupta, Rahul, and Sunita Sarawagi. 2011. *Joint training for open-domain extraction on the web: exploiting overlap when supervision is limited*. In: Proceedings of WSDM 2011, pp. 217-226.

Hearst, Marti A. 1992. *Automatic acquisition of hyponyms from large text corpora*. In: Proceedings of COLING 1992, pp. 539-545.

Hoffmann, Raphael, Congle Zhang, and Daniel S. Weld. 2010. *Learning 5000 relational extractors*. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 286-295.

Hoffmann, Raphael, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. 2011. *Knowledge-based weak supervision for information extraction of overlapping relations*. In: Proceedings of ACL 2011, pp. 541-550.

Ji, Heng, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2010. *Overview of the TAC 2010 knowledge base population track.* In: Proceedings of the Text Analytics Conference.

Ji, Heng, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2011. *Overview of the TAC 2011 knowledge base population track*. In Proceedings of the Text Analytics Conference.

Krause, Sebastian, Hong Li, Hans Uszkoreit, and Feiyu Xu. 2012. *Large-Scale learning of relation-extraction rules with distant supervision from the web*. In: ISWC 2012, pp. 263-278.

Mintz, Mike, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. *Distant supervision for relation extraction without labeled data*. In: Proceedings ACL-AFNLP 2009, pp. 1003-1011.

Min, Bonan, Ralph Grishman, Li Wan, Chang Wang, and David Gondek. 2013. *Distant Supervision for Relation Extraction with an Incomplete Knowledge Base*. In: Proceedings of NAACL-HLT 2013,pp. 777-782.

Min, Bonan, Xiang Li, Ralph Grishman, and Ang Sun. 2012. *New york university 2012 system for kbp slot filling*. In: Proceedings of TAC 2012.

Nguyen, Truc-Vien T., and Alessandro Moschitti. 2011. *Joint distant and direct supervision for relation extraction*. In: Proceedings of IJCNLP 2011, pp. 732-740.

Riedel, Sebastian, Limin Yao, and Andrew McCallum. 2010. *Modeling relations and their mentions without labeled text*. In: Machine Learning and Knowledge Discovery in Databases, 2010, pp. 148-163.

Riedel, Sebastian, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. 2013. *Relation Extraction with Matrix Factorization and Universal Schemas*. In: Proceedings of NAACL-HLT 2013, pp. 74-84.

Roth, Benjamin, and Dietrich Klakow. 2013. *Combining Generative and Discriminative Model Scores for Distant Supervision*. In: Proceedings of ACL 2013, pp. 24-29.

Surdeanu, Mihai, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. 2012. *Multi-instance multi-label learning for relation extraction*. In: Proceedings of EMNLP-CoNLL 2012, pp. 455-465.

Takamatsu, Shingo, Issei Sato, and Hiroshi Nakagawa. 2012. *Reducing wrong labels in distant supervision for relation extraction*. In: ACL 2012,pp. 721-729.

Wang, Chang, Aditya Kalyanpur, James Fan, Branimir K. Boguraev, and D. C. Gondek. 2012. *Relation extraction and scoring in DeepQA*. In: IBM Journal of Research and Development, 56(3.4), pp. 9-1.

Wang, Chang, James Fan, Aditya Kalyanpur, and David Gondek. 2011. *Relation extraction with relation topics*. In: Proceedings of EMNLP 2011, pp. 1426-1436.

Wright, John, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma. 2009. *Robust face recognition via sparse representation.* In: Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(2), 210-227

Wu, Fei, and Daniel S. Weld. 2007. Autonomously semantifying wikipedia. In: Proceedings of CIKM 2007,pp. 41-50.

Xu, Wei, Raphael Hoffmann Le Zhao, and Ralph Grishman. 2013. *Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction*. In: Proceedings of Proceedings of 2013, pp. 665-670.

Yang, Allen Y., Shankar S. Sastry, Arvind Ganesh, and Yi Ma. 2010. *Fast l^1-Minimization Algorithms and An Application in Robust Face Recognition: A Review*. In: Proceedings of ICIP 2010.

Yao, Limin, Sebastian Riedel, and Andrew McCallum. 2010. *Collective cross-document relation extraction* *without labelled data*. In: Proceedings of EMNLP 2010, pp. 1013-1023.

Zhang, Congle, Raphael Hoffmann, and Daniel S. Weld. 2012. *Ontological smoothing for relation extraction with minimal supervision*. In: Proceedings of AAAI 2012, pp. 157-163.

Zhang, Xingxing, Zhang, Jianwen, Zeng, Junyu, Yan, Jun, Chen, Zheng and Sui, Zhifang. 2013. *Towards Accurate Distant Supervision for Relational Facts Extraction*. In: Proceedings of ACL 2013, pp. 810-815.