A Graph-based Cross-lingual Projection Approach for Weakly Supervised Relation Extraction

Seokhwan Kim
Human Language Technology Dept.
Institute for Infocomm Research
Singapore 138632
kims@i2r.a-star.edu.sg

Gary Geunbae Lee
Dept. of Computer Science and Engineering
Pohang University of Science and Technology
Pohang, 790-784, Korea
gblee@postech.ac.kr

Abstract

Although researchers have conducted extensive studies on relation extraction in the last decade, supervised approaches are still limited because they require large amounts of training data to achieve high performances. To build a relation extractor without significant annotation effort, we can exploit cross-lingual annotation projection, which leverages parallel corpora as external resources for supervision. This paper proposes a novel graph-based projection approach and demonstrates the merits of it by using a Korean relation extraction system based on projected dataset from an English-Korean parallel corpus.

1 Introduction

Relation extraction aims to identify semantic relations of entities in a document. Although many supervised machine learning approaches have been successfully applied to relation extraction tasks (Zelenko et al., 2003; Kambhatla, 2004; Bunescu and Mooney, 2005; Zhang et al., 2006), applications of these approaches are still limited because they require a sufficient number of training examples to obtain good extraction results. Several datasets that provide manual annotations of semantic relationships are available from MUC (Grishman and Sundheim, 1996) and ACE (Doddington et al., 2004) projects, but these datasets contain labeled training examples in only a few major languages, including English, Chinese, and Arabic. Although these datasets encourage the development of relation extractors for these major languages, there are few labeled training samples for learning new systems in other languages, such as Korean. Because manual annotation of semantic relations for such *resource-poor languages* is very expensive, we instead consider weakly supervised learning techniques (Riloff and Jones, 1999; Agichtein and Gravano, 2000; Zhang, 2004; Chen et al., 2006) to learn the relation extractors without significant annotation efforts. But these techniques still face cost problems when preparing quality seed examples, which plays a crucial role in obtaining good extractions.

Recently, some researchers attempted to use external resources, such as treebank (Banko et al., 2007) and Wikipedia (Wu and Weld, 2010), that were not specially constructed for relation extraction instead of using task-specific training or seed examples. We previously proposed to leverage parallel corpora as a new kind of external resource for relation extraction (Kim et al., 2010). To obtain training examples in the resource-poor target language, this approach exploited a *cross-lingual annotation projection* by propagating annotations that were generated by a relation extraction system in a resource-rich source language. In this approach, projected annotations were determined in a single pass process by considering only alignments between entity candidates; we call this action *direct projection*.

In this paper, we propose a graph-based projection approach for weakly supervised relation extraction. This approach utilizes a graph that is constucted with both instance and context information and that is operated in an iterative manner. The goal of our graph-based approach is to improve the robustness of the extractor with respect to errors that are generated and accumulated by preprocessors.

Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 48–53,
Jeju, Republic of Korea, 8-14 July 2012.

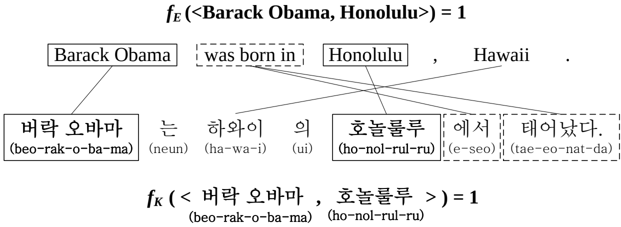

Figure 1: An example of annotation projection for relation extraction of a bitext in English and Korean

2 Cross-lingual Annotation Projection for Relation Extraction

Relation extraction can be considered to be a classification problem by the following classifier:

$$f\left(e^i, e^j\right) = \begin{cases} 1 & \text{if } e^i \text{ and } e^j \text{ have a relation,} \\ -1 & \text{otherwise.} \end{cases},$$

where e_i and e_j are entities in a sentence.

Cross-lingual annotation projection intends to learn an extractor f_t for good performance without significant effort toward building resources for a resource-poor target language L_t. To accomplish that goal, the method automatically creates a set of annotated text for f_t, utilizing a well-made extractor f_s for a resource-rich source language L_s and a parallel corpus of L_s and L_t. Figure 1 shows an example of annotation projection for relation extraction with a bi-text in L_t Korean and L_s English. Given an English sentence, an instance ⟨*Barack Obama, Honolulu*⟩ is extracted as positive. Then, its translational counterpart ⟨*beo-rak-o-ba-ma, ho-nol-rul-ru*⟩ in the Korean sentence also has a positive annotation by projection.

Early studies in cross-lingual annotation projection were accomplished for various natural language processing tasks (Yarowsky and Ngai, 2001; Yarowsky et al., 2001; Hwa et al., 2005; Zitouni and Florian, 2008; Pado and Lapata, 2009). These studies adopted a simple direct projection strategy that propagates the annotations in the source language sentences to word-aligned target sentences, and a target system can bootstrap from these projected annotations.

For relation extraction, the direct projection strategy can be formularized as follows: $f_t\left(e_t^i, e_t^j\right) = f_s\left(A(e_t^i), A(e_t^j)\right)$, where $A(e_t)$ is the aligned entity of e_t. However, these automatic annotations can be unreliable because of source text mis-classification and word alignment errors; thus, it can cause a critical falling-off in the annotation projection quality.

Although some noise reduction strategies for projecting semantic relations were proposed (Kim et al., 2010), the direct projection approach is still vulnerable to erroneous inputs generated by submodules. We note two main causes for this limitation: (1) the direct projection approach considers only alignments between entity candidates, and it does not consider any contextual information; and, (2) it is performed by a single pass process. To solve both of these problems at once, we propose a graph-based projection approach for relation extraction.

3 Graph Construction

The most crucial factor in the success of graph-based learning approaches is how to construct a graph that is appropriate for the target task. Das and Petrov (Das and Petrov, 2011) proposed a graph-based bilingual projection of part-of-speech tagging by considering the tagged words in the source language as labeled examples and connecting them to the unlabeled words in the target language, while referring to the word alignments. Graph construction for projecting semantic relationships is more complicated than part-of-speech tagging because the unit instance of projection is a pair of entities and not a word or morpheme that is equivalent to the alignment unit.

3.1 Graph Vertices

To construct a graph for a relation projection, we define two types of vertices: instance vertices V and context vertices U.

Instance vertices are defined for all pairs of entity candidates in the source and target languages. Each instance vertex has a soft label vector $Y = [\; y^+ \quad y^- \;]$, which contains the probabilities that the instance is positive or negative, respectively. The larger the y^+ value, the more likely the instance has a semantic relationship. The initial label values of an instance vertex $v_s^{ij} \in V_s$ for the instance $\left\langle e_s^i, e_s^j \right\rangle$ in the source language are assigned based on the confidence score of the extractor f_s. With respect to the target language, every instance vertex $v_t^{ij} \in V_t$ has the same initial values of 0.5 in both y^+ and y^-.

The other type of vertices, context vertices, are used for identifying relation descriptors that are contextual subtexts that represent semantic relationships of the positive instances. Because the characteristics of these descriptive contexts vary depending on the language, context vertices should be defined to be language-specific. In the case of English, we define the context vertex for each trigram that is located between a given entity pair that is semantically related. If the context vertices U_s for the source language sentences are defined, then the units of context in the target language can also be created based on the word alignments. The aligned counterpart of each source language context vertex is used for generating a context vertex $u_t^i \in U_t$ in the target language. Each context vertex $u_s \in U_s$ and $u_t \in U_t$ also has y^+ and y^-, which represent how likely the context is to denote semantic relationships. The probability values for all of the context vertices in both of the languages are initially assigned to $y^+ = y^- = 0.5$.

3.2 Edge Weights

The graph for our graph-based projection is constructed by connecting related vertex pairs by weighted edges. If a given pair of vertices is likely to have the same label, then the edge connecting these vertices should have a large weight value.

We define three types of edges according to combinations of connected vertices. The first type of edges consists of connections between an instance vertex and a context vertex in the same language. For a pair of an instance vertex $v^{i,j}$ and a context vertex u^k, these vertices are connected if the context sequence of $v^{i,j}$ contains u^k as a subsequence. If v^{ij} is matched to u^k, the edge weight $w\left(v^{i,j}, u^k\right))$ is assigned to 1. Otherwise, it should be 0.

Another edge category is for the pairs of context vertices in a language. Because each context vertex is considered to be an n-gram pattern in our work, the weight value for each edge of this type represents the pattern similarity between two context vertices. The edge weight $w(u^k, u^l)$ is computed by Jaccard's coefficient between u^k and u^l.

While the previous two categories of edges are concerned with monolingual connections, the other type addresses bilingual alignments of context vertices between the source language and the target language. We define the weight for a bilingual edge connecting u_s^k and u_t^l as the relative frequency of alignments, as follows:

$$w(u_s^k, u_t^l) = \text{count}\left(u_s^k, u_t^l\right) / \sum_{u_t^m} \text{count}\left(u_s^k, u_t^m\right),$$

where $\text{count}\,(u_s, u_t)$ is the number of alignments between u_s and u_t across the whole parallel corpus.

4 Label Propagation

To induce labels for all of the unlabeled vertices on the graph constructed in Section 3, we utilize the label propagation algorithm (Zhu and Ghahramani, 2002), which is a graph-based semi-supervised learning algorithm.

First, we construct an $n \times n$ matrix T that represents transition probabilities for all of the vertex pairs. After assigning all of the values on the matrix, we normalize the matrix for each row, to make the element values be probabilities. The other input to the algorithm is an $n \times 2$ matrix Y, which indicates the probabilities of whether a given vertex v_i is positive or not. The matrix T and Y are initialized by the values described in Section 3.

For the input matrices T and Y, label propagation is performed by multiplying the two matrices, to update the Y matrix. This multiplication is repeated until Y converges or until the number of iterations exceeds a specific number. The Y matrix, after finishing its iterations, is considered to be the result of the algorithm.

5 Implementation

To demonstrate the effectiveness of the graph-based projection approach for relation extraction, we developed a Korean relation extraction system that was trained with projected annotations from English resources. We used an English-Korean parallel corpus [1] that contains 266,892 bi-sentence pairs in English and Korean. We obtained 155,409 positive instances from the English sentences using an off-the-shelf relation extraction system, ReVerb [2] (Fader et al., 2011).

[1]The parallel corpus collected is available in our website: http://isoft.postech.ac.kr/~megaup/acl/datasets

[2]http://reverb.cs.washington.edu/

Table 1: Comparison between direct and graph-based projection approaches to extract semantic relationships for four relation types

Type	Direct			Graph-based		
	P	R	F	P	R	F
Acquisition	51.6	87.7	64.9	55.3	91.2	68.9
Birthplace	69.8	84.5	76.4	73.8	87.3	80.0
Inventor Of	62.4	85.3	72.1	66.3	89.7	76.3
Won Prize	73.3	80.5	76.7	76.4	82.9	79.5
Total	63.9	84.2	72.7	67.7	87.4	76.3

The English sentence annotations in the parallel corpus were then propagated into the corresponding Korean sentences. We used the GIZA++ software [3] (Och and Ney, 2003) to obtain the word alignments for each bi-sentence in the parallel corpus. The graph-based projection was performed by the Junto toolkit [4] with the maximum number of iterations of 10 for each execution.

Projected instances were utilized as training examples to learn the Korean relation extractor. We built a tree kernel-based support vector machine model using SVM-Light [5] (Joachims, 1998) and Tree Kernel tools [6] (Moschitti, 2006). In our model, we adopted the subtree kernel method for the shortest path dependency kernel (Bunescu and Mooney, 2005).

6 Evaluation

The experiments were performed on the manually annotated Korean test dataset. The dataset was built following the approach of Bunescu and Mooney (Bunescu and Mooney, 2007). The dataset consists of 500 sentences for four relation types: Acquisition, Birthplace, Inventor of, and Won Prize. Of these, 278 sentences were annotated as positive instances.

The first experiment aimed to compare two systems constructed by the direct projection (Kim et al., 2010) and graph-based projection approach. Table 1 shows the performances of the relation extraction of the two systems. The graph-based system achieved better performances in precision and recall than the system with direct projection for all of the four relation types. It outperformed the baseline system by an F-measure of 3.63.

[3]http://code.google.com/p/giza-pp/

[4]http://code.google.com/p/junto/

[5]http://svmlight.joachims.org/

[6]http://disi.unitn.it/ moschitt/Tree-Kernel.htm

Table 2: Comparisons of our projection approach to heuristic and Wikipedia-based approaches

Approach	P	R	F
Heuristic-based	92.31	17.27	29.09
Wikipedia-based	66.67	66.91	66.79
Projection-based	67.69	87.41	76.30

To demonstrate the merits of our work against other approaches based on monolingual external resources, we performed comparisons with the following two baselines: heuristic-based (Banko et al., 2007) and Wikipedia-based approaches (Wu and Weld, 2010). The heuristic-based baseline was built on the Sejong treebank corpus (Kim, 2006) and the Wikipedia-based baseline used Korean Wikipedia articles [7]. Table 2 compares the performances of the two baseline systems and our method. Our proposed projection-based approach obtained better performance than the other systems. It outperformed the heuristic-based system by 47.21 and the Wikipedia-based system by 9.51 in the F-measure.

7 Conclusions

This paper presented a novel graph-based projection approach for relation extraction. Our approach performed a label propagation algorithm on a proposed graph that represented the instance and context features of both the source and target languages. The feasibility of our approach was demonstrated by our Korean relation extraction system. Experimental results show that our graph-based projection helped to improve the performance of the cross-lingual annotation projection of the semantic relations, and our system outperforms the other systems, which incorporate monolingual external resources.

In this work, we operated the graph-based projection under very restricted conditions, because of high complexity of the algorithm. For future work, we plan to relieve the complexity problem for dealing with more expanded graph structure to improve the performance of our proposed approach.

[7]We used the Korean Wikipedia database dump as of June 2011.

Acknowledgments

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2012-(H0301-12-3001)) supervised by the NIPA(National IT Industry Promotion Agency) and Industrial Strategic technology development program, 10035252, development of dialog-based spontaneous speech interface technology on mobile platform, funded by the Ministry of Knowledge Economy(MKE, Korea).

References

E. Agichtein and L. Gravano. 2000. Snowball: Extracting relations from large plain-text collections. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 85–94.

M. Banko, M. J Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. 2007. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2670–2676.

R. Bunescu and R. Mooney. 2005. A shortest path dependency kernel for relation extraction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 724–731.

R. Bunescu and R. Mooney. 2007. Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th annual meeting of the Association for Computational Linguistics*, volume 45, pages 576–583.

J. Chen, D. Ji, C. L Tan, and Z. Niu. 2006. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 129–136.

D. Das and S. Petrov. 2011. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 600–609.

G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. 2004. The automatic content extraction (ACE) program–tasks, data, and evaluation. In *Proceedings of LREC*, volume 4, pages 837–840.

A. Fader, S. Soderland, and O. Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545.

R. Grishman and B. Sundheim. 1996. Message understanding conference-6: A brief history. In *Proceedings of the 16th conference on Computational linguistics*, volume 1, pages 466–471.

R. Hwa, P. Resnik, A. Weinberg, C. Cabezas, and O. Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11(3):311–325.

T. Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the European Conference on Machine Learning*, pages 137–142.

N. Kambhatla. 2004. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, pages 22–25.

S. Kim, M. Jeong, J. Lee, and G. G Lee. 2010. A cross-lingual annotation projection approach for relation detection. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 564–571.

H. Kim. 2006. Korean national corpus in the 21st century sejong project. In *Proceedings of the 13th NIJL International Symposium*, pages 49–54.

A. Moschitti. 2006. Making tree kernels practical for natural language learning. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, volume 6, pages 113–120.

F. J Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

S. Pado and M. Lapata. 2009. Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, 36(1):307–340.

E. Riloff and R. Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the National Conference on Artificial Intelligence*, pages 474–479.

F. Wu and D. Weld. 2010. Open information extraction using wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127.

D. Yarowsky and G. Ngai. 2001. Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 1–8.

D. Yarowsky, G. Ngai, and R. Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the*

First International Conference on Human Language Technology Research, pages 1–8.

D. Zelenko, C. Aone, and A. Richardella. 2003. Kernel methods for relation extraction. *The Journal of Machine Learning Research*, 3:1083–1106.

M. Zhang, J. Zhang, J. Su, and G. Zhou. 2006. A composite kernel to extract relations between entities with both flat and structured features. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 825–832.

Z. Zhang. 2004. Weakly-supervised relation classification for information extraction. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 581–588.

X. Zhu and Z. Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation. *School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CALD-02-107*.

I. Zitouni and R. Florian. 2008. Mention detection crossing the language barrier. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 600–609.