A GRAPH-BASED CROSS-LINGUAL PROJECTION APPROACH FOR SPOKEN LANGUAGE UNDERSTANDING PORTABILITY TO A NEW LANGUAGE

Seokhwan Kim

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

ABSTRACT

The portability of spoken language understanding to a new language can be improved by the results of automatic translation. However, the translation errors can cause the falling-off in the quality of the target language system. This paper proposes a graph-based projection approach to improve the robustness against the translation errors in cross-lingual spoken language understanding. The experimental results show that our proposed approach can significantly improve the performances of the task in a new language.

Index Terms— Spoken Dialogue Systems, Spoken Language Understanding, Language Portability, Statistical Machine Translation

1. INTRODUCTION

Statistical approaches to spoken language understanding (SLU) have been shown to reduce the development time and cost of the systems in comparison with hand-written grammar-based approaches [1, 2, 3, 4, 5]. However, they also require a sufficient number of training examples to obtain good results; thus, even if there exists a well developed dataset in a certain language, we need to collect, transcribe, and annotate a separate dataset manually to build an SLU system for a different language.

Recently, some researchers attempted to use statistical machine translation (SMT) technologies to improve the portability of SLU to a new language [6, 7, 8, 9]. The key to these cross-lingual SLU approaches is to transfer the semantic knowledge from the existing resources in the source language L_s to the target language L_t based on the SMT results between L_s and L_t. The major obstacle to the success of cross-lingual SLU approaches is due to the imperfectness of current SMT technologies. Even with the state-of-the-art SMT systems, automatically translated utterances tend to include a certain number of translation errors.

Although some noise reduction strategies for cross-lingual SLU were proposed, these studies have focused on filtering out or correcting the noisy translations as post-processing after projecting the semantic knowledge [8, 9]. These noise reduction strategies are performed in a single pass process by considering only the translations for each utterance independently.

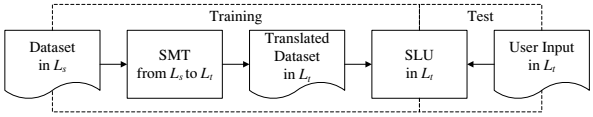
In this paper, we propose a graph-based projection approach for cross-lingual SLU. This approach utilizes a graph that is constructed with whole dataset and that is operated in an iterative manner to improve the robustness to the noisy translations. An early study in graph-based projection was accomplished for cross-lingual part-of-speech tagging [10]. While it is working with the manually aligned parallel corpus at the sentence level, our approach is based on the results of automatic translation which are supposed to be much noisier than the manual one. In addition to that, our proposed approach aims to project not only the word-level semantics, but also the utterance-level categories, which that leads to more complicated graph structure than previous work.

We present an overview of cross-lingual SLU using SMT in Section 2. We describe our proposed approach to cross-lingual SLU based on a graph-based learning algorithm in Section 3, and present details about the implementation of the English-Korean SLU systems developed based on the graph-based approach in Section 4. We report the evaluation result of the system in Section 5, and we conclude this paper in Section 6.

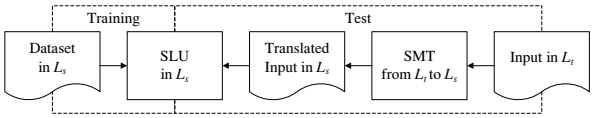
2. CROSS-LINGUAL SLU USING SMT

Cross-lingual SLU can be performed in either of or combination of two major strategies: TrainOnTarget and TestOnSource [7]. In TrainOnTarget approach, the training dataset is translated from L_s to L_t in the beginning of the training phase. Then, this translated dataset is used to train the SLU model in L_t. In the execution phase, the test utterances in L_t are inputted to the trained model directly (Figure 1(a)).

On the other hand, TestOnSource approach utilizes the SLU model L_s trained on the manually built dataset. Then, the automatic translation from L_t to L_s is performed in the ex-



(a) TrainOnTarget



(b) TestOnSource

Fig. 1. Cross-lingual SLU strategies

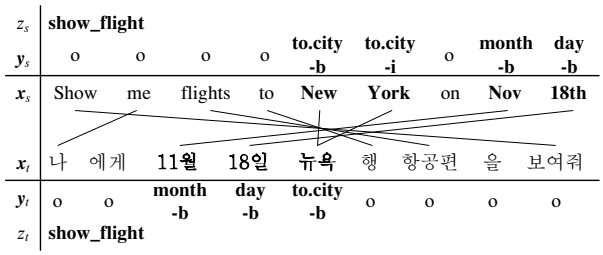


Fig. 2. An example of cross-lingual projection for SLU in English and Korean

ecution phase to generate the input to the model (Figure 1(b)). Previous work showed that TestOnSource achieved better performance than TrainOnTarget with no noisy reduction strategy [7, 8]. However, the room for improvement with TestOnSource is relatively limited, because only a translation pair for the single input utterance can be available in the execution phase.

Thus, our proposed approach focuses on improving TrainOnTarget strategy by considering not only a single instance, but also the whole dataset in the offline training phase. In TrainOnTarget approach, the annotations on L_s utterances should be projected onto their corresponding translations in L_t. The annotations for SLU are usually divided into two main components: named entity (NE) for each word and dialog act (DA) for each utterance. For a given word sequence in an utterance $\boldsymbol{x} = \{x_1, \cdots, x_n\}$, NE annotations can be represented as an NE tag sequence $\boldsymbol{y} = \{y_1, \cdots, y_n\}$; and the DA annotation should be a class variable z. Both NE and DA annotations are projected from L_s to L_t based on the word alignments generated during the SMT process, as shown in the example in Figure 2.

3. GRAPH-BASED PROJECTION

The simplest way of projection is to propagate the annotations by considering only word alignments themselves; we call this action *direct projection*. For a given utterance $\boldsymbol{x}_s = \{x_s^1, \cdots, x_s^n\}$ in L_s and its translation $\boldsymbol{x}_t = \{x_t^1, \cdots, x_t^m\}$ in L_t, the projected annotations are determined with direct projection as follows:

$$\begin{aligned}\boldsymbol{y}(\boldsymbol{x}_t) &= \{NE(x_t^1), \cdots, NE(x_t^m)\} \\ &= \{NE(A_s(x_t^1)), \cdots, NE(A_s(x_t^m))\}), \\ z(X_t) &= DA(\{x_t^1, \cdots, x_t^m\}) \\ &= DA(\{A_s(x_t^1), \cdots, A_s(x_t^m)\}),\end{aligned}$$

where $NE(x)$ is the NE annotation for a word x, $DA(\boldsymbol{x})$ is the DA annotation for a word sequence $\boldsymbol{x}$, and $A_s(x_t)$ is the aligned word of x_t in $\boldsymbol{x}_s$. However, the propagated annotations with direct projection can be unreliable when there are erroneous inputs generated by automatic translation and word alignment. We note two main causes for this limitation: (1) the direct projection approach considers only the translation for each single utterance; 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 cross-lingual SLU. Our proposed approach is performed in two phases: graph construction and label propagation.

3.1. 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. Since we are aiming to project two different types of annotations: NE and DA, the graph structure should be defined separately for each task.

3.1.1. Graph for NE projection

To construct a graph for NE projection, we define n-gram nodes for both languages and connect them with weighted edges. First, the monolingual parts of the graph for each language are constructed with the structure defined in Subramanya et al. [11] that is for semi-supervised learning of monolingual tagging. The nodes $V = \{v_1, \cdots, v_n\}$ are defined for all trigrams in the dataset, then the contextual similarities for all node pairs are computed as follows:

$$w(v_i, v_j) = sim_{cosine}(f(v_i), f(v_j)) = \frac{f(v_i) \cdot f(v_j)}{|f(v_i)||f(v_j)|},$$

where $f(v)$ is the feature vector of the node v, that defined in [11]. With the similarities for all node pairs, a nearest neighbor graph is constructed by assigning the edge weights to the computed similarity values for the n most similar nodes of a given node and to 0 for other nodes.

After both monolingual subgraphs L_s and L_t are constructed, bilingual connections are established with the edge weights defined as follows:

$$w(v_s^k, v_t^l) = \frac{\text{count}\left(v_s^k, v_t^l\right)}{\sum_{v_t^m} \text{count}\left(v_s^k, v_t^m\right)},$$

where v_s is a node in L_s, v_t is a node in L_t, and $\mathrm{count}\,(v_s, v_t)$ is the number of alignments between v_s and v_t across the whole translated dataset.

Since this graph is defined for propagating the NE labels, each node has a label distribution vector where its length is same to the number of NE labels. The probabilities that the node belongs to corresponding NE labels are encoded in this vector. The initial values for label vectors are assigned based on the manual annotations of NE in L_s only.

3.1.2. Graph for DA projection

The unit instance of DA projection is an utterance and not a word that is equivalent to the alignment unit. Thus, we define the utterance nodes $U = \{u_1, \cdots, u_m\}$ in addtion to the graph for NE projection. An utterance node u_i corresponds to each utterance and is connected to the trigram nodes in the same language only. The edge between u_i and v_j has a binary weight value as follows:

$$w\,(u_i, v_j) = \begin{cases} 1 & \text{if } v_j \text{ in } u_i, \\ 0 & \text{otherwise.} \end{cases} .$$

The utterance nodes in L_s can have the initial values in this graph for DA projection.

Figure 3 shows the comparison of graph structures between NE and DA projections. Solid lines and dotted lines mean monolingual and bilingual connections respectively; and Gray-colored circles and uncolored circles mean labeled and unlabeled nodes at the initial step respectively.

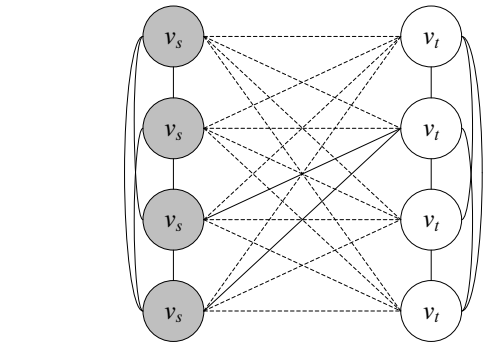
3.2. Label Propagation

To induce labels for all of the unlabeled nodes on the graph constructed in Section 3.1, we utilize the label propagation algorithm [12], which is a graph-based semi-supervised learning algorithm (Figure 4).

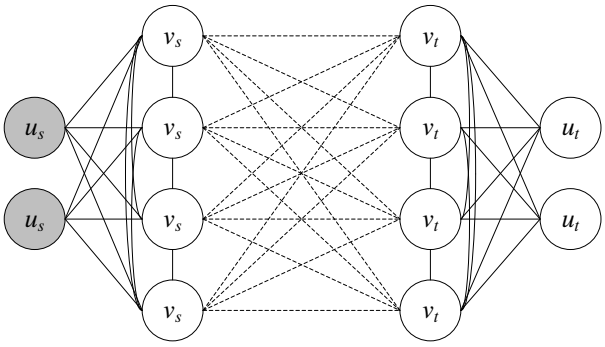
First, we construct an $n \times n$ matrix T that represents transition probabilities for all of the node pairs. Each element T_{ij} represents the probability of propagating a label from node v_j to another node v_i and is computed based on w_{ij}, which is defined in Section 3.1. After assigning all 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 $n \times m$ matrix Y. The value of Y_{ij} is the probability that a given node v_i belongs to the j-th label. The matrix Y is initialized by the values also described in Section 3.1.

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.



(a) Graph for NE projection



(b) Graph for DA projection

Fig. 3. Graph Structures for Cross-lingual SLU

4. IMPLEMENTATION

To demonstrate the effectiveness of the graph-based projection approach for cross-lingual SLU, we developed two SLU systems in English and Korean on tourist information dialogues. The dataset consists of 3,351 pairs of bi-utterances that are translational counterparts to each other in English and Korean. All utterances in both languages were manually annotated with 30 DA classes and 30 NE classes which are defined for the target domain.

From these manually translated and annotated references, we generated the automatic translated datasets in both directions, English to Korean and Korean to English. To perform these translations, we used an SMT system trained on Infinity corpus [13] with Moses [1] [14] and SRILM [2] [15] toolkits. This system achieved 21.01% in BLEU score from English to Korean and 28.36% in BLUE score from Korean to English.

We constructed the graphs for our proposed approach with the SMT results including translated utterances and word alignments. Table 1 shows the sizes of graphs generated for four combinations of subtasks and projection directions. In these graphs, the number of nearest neighbors for monolin-

[1] http://www.statmt.org/moses/

[2] http://www.speech.sri.com/projects/srilm/

for v_i **in** V **do**
  for v_j **in** V **do**
    $T_{ij} \leftarrow w_{ij} / \left(\sum_k w_{kj}\right)$ ▷ Initialize T
for v_i **in** V **do**
  for v_j **in** V **do**
    $T_{ij} \leftarrow T_{ij} / \left(\sum_k T_{ik}\right)$ ▷ Normalize T
for v_i **in** V **do**
  $Y_i^0 \leftarrow \left[Y_{i1}\left(v_i\right), \cdots, Y_{im}\left(v_i\right)\right]$ ▷ Initialize Y
$t \leftarrow 0$
repeat
  $Y^{t+1} \leftarrow TY^t$ ▷ Update Y
  $t \leftarrow t + 1$
until $t \geq$ maxiter **or** Y converges
return Y^t

Fig. 4. Label Propagation Algorithm

Table 1. Statistics of the graphs for cross-lingual projection in English and Korean

	Korean→English		English→Korean	
	# of nodes	# of edges	# of nodes	# of edges
NE	38,818	60,733	72,484	117,545
DA	42,610	137,088	79,480	259,232

gual connections was set to five nodes. Then, we performed projections with Junto label propagation toolkit [3] [16].

These translated utterances with projected annotations were used for training the SLU models in target language. We used maximum entropy (ME) and conditional random fields (CRF) models for DA identification and NE recognition, respectively. Both types of models were trained with FastCRF [4] toolkit with word n-gram features.

5. EVALUATION

The experiments were performed with the following four different strategies: *Supervised* model was trained with manual annotations on the monolingual utterances in L_t only; *TestOnSource* used the monolingual supervised model in L_s to predict the semantics of the translated utterances from L_t to L_s; *Direct* approach was based on TrainOnTarget strategy without any noise reduction methods; and *Graph-based* strategy trained the model with our proposed approach. All the evaluations were done in five-fold cross validation to the manual annotations on L_t utterances with the metrics of precision, recall, and F-measure for NE recognition and accuracy for DA identification.

Table 2 compares the performances of NE recognition with four approaches. As pervious work reported, TrainOnTarget strategy with direct projection failed to obtain better performances than TestOnSource. However, our proposed graph-based projection approach achieved significant performance improvements in both precision and recall. It outperformed the direct-projection model by F-measures of 25.9 in English and 29.7 in Korean; and it also obtained higher performances than TestOnSource model by F-measures of 20.6 in English and 22.4 in Korean.

[3]https://github.com/parthatalukdar/junto

[4]https://github.com/minwoo/fastCRF

Table 2. Comparison of NE recognition performances among the cross-lingual SLU strategies

	Korean→English			English→Korean		
	P	R	F	P	R	F
Supervised	97.6	95.4	96.4	97.1	96.9	97.0
TestOnSource	45.2	16.4	24.0	63.8	19.9	30.3
Direct	43.1	11.9	18.7	50.9	14.8	23.0
Graph-based	50.7	39.8	44.6	67.2	43.4	52.7

Table 3. Comparison of DA identification performances among the cross-lingual SLU strategies

	Accuracy (%)	
	Korean→English	English→Korean
Supervised	87.7	83.3
TestOnSource	58.9	70.2
Direct	56.5	69.6
Graph-based	63.5	74.3

The results of DA identification also show the similar aspect to the NE recognition results (Table 3). Our proposed approach achieved better accuracy than the direct-projection model by 7.0% in English and 4.7% in Korean; and than the TestOnSource model by 4.6% in English and 4.1% in Korean.

6. CONCLUSION

This paper presented a graph-based projection approach for cross-lingual SLU using SMT. Our approach performed a label propagation algorithm on a proposed graph that was defined with the translations for all over the dataset. The feasibility of our approach was demonstrated by English and Korean SLU models. Experimental results show that our graph-based projection helped to improve the performances of the cross-lingual SLU than previous approaches.

In this work, we operated the graph-based projection only in a single direction from the manually labeled annotations in L_s to the unlabeled instances in L_t. For future work, we plan to investigate the way of bi-directional projection to complete the partially labeled annotations or to improve the quality of existing datasets.

7. REFERENCES

[1] E. Levin and R. Pieraccini, "Chronus, the next generation," in *Proceedings of the DARPA Speech and Natural Language Workshop*, 1995, pp. 269–271.

[2] R. Schwartz, S. Miller, D. Stallard, and J. Makhoul, "Language understanding using hidden understanding models," in *Proceedings of the fourth International Conference on Spoken Language Processing $ICSLP$*, 1996, vol. 2, pp. 997–1000.

[3] Y. He and S. Young, "Spoken language understanding using the hidden vector state model," *Speech Communication*, vol. 48, no. 3, pp. 262–275, 2006.

[4] Y.Y. Wang and A. Acero, "Discriminative models for spoken language understanding," in *Proceedings of the Ninth International Conference on Spoken Language Processing $ICSLP$*, 2006, pp. 2426–2429.

[5] S. Hahn, M. Dinarelli, C. Raymond, F. Lefèvre, P. Lehnen, R. De Mori, A. Moschitti, H. Ney, and G. Riccardi, "Comparing stochastic approaches to spoken language understanding in multiple languages," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1569–1583, 2011.

[6] C. Servan, N. Camelin, C. Raymond, F. Béchet, and R. De Mori, "On the use of machine translation for spoken language understanding portability," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 5330–5333.

[7] F. Lefèvre, F. Mairesse, and S. Young, "Cross-lingual spoken language understanding from unaligned data using discriminative classification models and machine translation," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, 2010, pp. 78–81.

[8] B. Jabaian, L. Besacier, and F. Lefèvre, "Combination of stochastic understanding and machine translation systems for language portability of dialogue systems," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 5612–5615.

[9] T. Misu, E. Mizukami, H. Kashioka, S. Nakamura, and H. Li, "A bootstrapping approach for slu portability to a new language by inducting unannotated user queries," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 4961–4964.

[10] D. Das and S. Petrov, "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*, 2011, pp. 600–609.

[11] A. Subramanya, S. Petrov, and F. Pereira, "Efficient graph-based semi-supervised learning of structured tagging models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 167–176.

[12] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Tech. Rep., CMU-CALD-02-107, Carnegie Mellon University, 2002.

[13] J. Lee, S. Lee, H. Noh, K. Lee, and G.G. Lee, "Iteratively constrained selection of word alignment links using knowledge and statistics," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1120–1130, 2011.

[14] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al., "Moses: Open source toolkit for statistical machine translation," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2007, vol. 45, p. 2.

[15] A. Stolcke et al., "Srilm-an extensible language modeling toolkit," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 2002, vol. 2, pp. 901–904.

[16] P.P. Talukdar and F. Pereira, "Experiments in graph-based semi-supervised learning methods for class-instance acquisition," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2010, pp. 1473–1481.