Semantic Roles for String to Tree Machine Translation

Marzieh Bazrafshan and **Daniel Gildea**
Department of Computer Science
University of Rochester
Rochester, NY 14627

Abstract

We experiment with adding semantic role information to a string-to-tree machine translation system based on the rule extraction procedure of Galley et al. (2004). We compare methods based on augmenting the set of nonterminals by adding semantic role labels, and altering the rule extraction process to produce a separate set of rules for each predicate that encompass its entire predicate-argument structure. Our results demonstrate that the second approach is effective in increasing the quality of translations.

1 Introduction

Statistical machine translation (SMT) has made considerable advances in using syntactic properties of languages in both the training and the decoding of translation systems. Over the past few years, many researchers have started to realize that incorporating semantic features of languages can also be effective in increasing the quality of translations, as they can model relationships that often are not derivable from syntactic structures.

Wu and Fung (2009) demonstrated the promise of using features based on semantic predicate-argument structure in machine translation, using these feature to re-rank machine translation output. In general, re-ranking approaches are limited by the set of translation hypotheses, leading to a desire to incorporate semantic features into the translation model used during MT decoding.

Liu and Gildea (2010) introduced two types of semantic features for tree-to-string machine translation. These features model the reorderings and deletions of the semantic roles in the source sentence during decoding. They showed that addition of these semantic features helps improve the quality of translations. Since tree-to-string systems are

trained on parse trees, they are constrained by the tree structures and are generally outperformed by string-to-tree systems.

Xiong et al. (2012) integrated two discriminative feature-based models into a phrase-based SMT system, which used the semantic predicate-argument structure of the source language. Their first model defined features based on the context of a verbal predicate, to predict the target translation for that verb. Their second model predicted the reordering direction between a predicate and its arguments from the source to the target sentence.

Wu et al. (2010) use a head-driven phrase structure grammar (HPSG) parser to add semantic representations to their translation rules.

In this paper, we use semantic role labels to enrich a string-to-tree translation system, and show that this approach can increase the BLEU (Papineni et al., 2002) score of the translations. We extract GHKM-style (Galley et al., 2004) translation rules from training data where the target side has been parsed and labeled with semantic roles. Our general method of adding information to the syntactic tree is similar to the "tree grafting" approach of Baker et al. (2010), although we focus on predicate-argument structure, rather than named entity tags and modality. We modify the rule extraction procedure of Galley et al. (2004) to produce rules representing the overall predicate-argument structure of each verb, allowing us to model alternations in the mapping from syntax to semantics of the type described by Levin (1993).

2 Semantic Roles for String-to-Tree Translation

2.1 Semantic Role Labeling

Semantic Role Labeling (SRL) is the task of identifying the arguments of the predicates in a sentence, and classifying them into different argument labels. Semantic roles can provide a level

of understanding that cannot be derived from syntactic analysis of a sentence. For example, in sentences "*Ali opened the door.*" and "*The door opened*", the word *door* has two different syntactic roles but only one semantic role in the two sentences.

Semantic arguments can be classified into core and non-core arguments (Palmer et al., 2010). Core arguments are necessary for understanding the sentence. Non-core arguments add more information about the predicate but are not essential.

Automatic semantic role labelers have been developed by training classifiers on hand annotated data (Gildea and Jurafsky, 2000; Srikumar and Roth, 2011; Toutanova et al., 2005; Fürstenau and Lapata, 2012). State-of-the-art semantic role labelers can predict the labels with accuracies of around 90%.

2.2 String-to-Tree Translation

We adopt the GHKM framework of Galley et al. (2004) using the parses produced by the split-merge parser of Petrov et al. (2006) as the English trees. As shown by Wang et al. (2010), the refined nonterminals produced by the split-merge method can aid machine translation. Furthermore, in all of our experiments, we exclude unary rules during extraction by ensuring that no rules will have the same span in the source side (Chung et al., 2011).

2.3 Using Semantic Role Labels in SMT

To incorporate semantic information into a string-to-tree SMT system, we tried two approaches:

- Using semantically enriched GHKM rules, and

- Extracting semantic rules separately from the regular GHKM rules, and adding a new feature for distinguishing the semantic rules.

The next two sections will explain these two methods in detail.

2.4 Semantically Enriched Rules (Method 1)

In this method, we tag the target trees in the training corpus with semantic role labels, and extract the translation rules from the tagged corpus. Since the SCFG rule extraction methods do not assume any specific set of non-terminals for the target parse trees, we can attach the semantic roles of each constituent to its label in the tree, and use
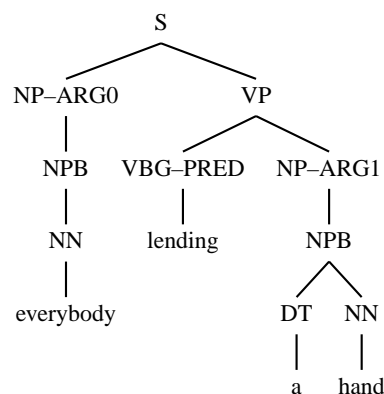

Figure 1: A target tree after inserting semantic roles. "Lending" is the predicate, "everybody" is argument 0, and "a hand" is argument 1 for the predicate.
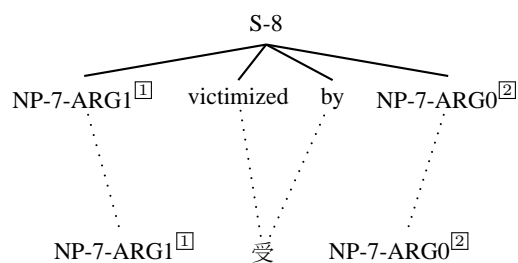

Figure 2: A complete semantic rule.

these new labels for rule extraction. We only label the core arguments of each predicate, to make sure that the rules are not too specific to the training data. We attach each semantic label to the root of the subtree that it is labeling. Figure 1 shows an example target tree after attaching the semantic roles. We then run a GHKM rule extractor on the labeled training corpus and use the semantically enriched rules with a syntax-based decoder.

2.5 Complete Semantic Rules with Added Feature (Method 2)

This approach uses the semantic role labels to extract a set of special translation rules, that on the target side form the smallest tree fragments in which one predicate and all of its core arguments are present. These rules model the complete semantic structure of each predicate, and are used by the decoder in addition to the normal GHKM rules, which are extracted separately.

Starting by semantic role labeling the target parse trees, we modify the GHKM component of the system to extract a semantic rule for each predicate. We define *labels_p* as the set of semantic role labels related to predicate p. That includes all

	Number of rules	
	dev	test
Baseline	1292175	1300589
Method 1	1340314	1349070
Method 2	1416491	1426159

Table 1: The number of the translation rules used by the three experimented methods

of the labels of the arguments of p, and the label of p itself. Then we add the following condition to the definition of the "frontier node" defined in Galley et al. (2004):

A frontier node must have either all or none of the semantic role labels from labels_p in its descendants in the tree.

Adding this new condition, we extract one semantic rule for each predicate, and for that rule we discard the labels related to the other predicates. This semantic rule will then have on its target side, the smallest tree fragment that contains all of the arguments of predicate p and the predicate itself.

Figure 2 depicts an example of a complete semantic rule. Numbers following grammatical categories (for example, S-8 at the root) are the refined nonterminals produced by the split-merge parser. In general, the tree side of the rule may extend below the nodes with semantic role labels because of the general constraint on frontier nodes that they must have a continuous span in the source (Chinese) side. Also, the internal nodes of the rules (such as a node with PRED label in Figure 2) are removed because they are not used in decoding.

We also extract the regular GHKM rules using the original definition of the frontier nodes, and add the semantic rules to them. To differentiate the semantic rules from the non-semantic ones, we add a new binary feature that is set to 1 for the semantic rules and to 0 for the rest of the rules.

3 Experiments

Semantic role labeling was done using the Prop-Bank standard (Palmer et al., 2005). Our labeler uses a maximum entropy classifier and for identification and classification of semantic roles, and has a persicion of 90% and a recall of 88%. The features used for training the labeler are a subset of the features used by Gildea and Jurafsky (2000), Xue and Palmer (2004), and Pradhan et al. (2004).

The string-to-tree training data that we used is a Chinese to English parallel corpus that contains more than 250K sentence pairs, which consist of 6.3M English words. The corpus was drawn from the newswire texts available from LDC.[1] We used a 392-sentence development set with four references for parameter tuning, and a 428-sentence test set with four references for testing. They are drawn from the newswire portion of NIST evaluation (2004, 2005, 2006). The development set and the test set only had sentences with less than 30 words for decoding speed. A set of nine standard features, which include globally normalized count of rules, lexical weighting (Koehn et al., 2003), length penalty, and number of rules used, was used for the experiments. In all of our experiments, we used the split-merge parsing method of Petrov et al. on the training corpus, and mapped the semantic roles from the original trees to the result of the split-merge parser. We used a syntax-based decoder with Earley parsing and cube pruning (Chiang, 2007). We used the Minimum Error Rate Training (Och, 2003) to tune the decoding parameters for the development set and tested the best weights that were found on the test set.

We ran three sets of experiments: Baseline experiments, where we did not do any semantic role labeling prior to rule extraction and only extracted regular GHKM rules, experiments with our method of Section 2.4 (Method 1), and a set of experiments with our method of Section 2.5 (Method 2).

Table 1 contains the numbers of the GHKM translation rules used by our three method. The rules were filtered by the development and the test to increase the decoding speed. The increases in the number of rules were expected, but they were not big enough to significantly change the performance of the decoder.

3.1 Results

For every set of experiments, we ran MERT on the development set with 8 different starting weight vectors picked randomly. For Method 2 we added a new random weight for the new feature. We then tested the system on the test set, using for each experiment the weight vector from the iteration of MERT with the maximum BLEU score on the development set. Table 3 shows the BLEU scores that we found on the test set, and their corresponding scores on the development set.

---

[1] We randomly sampled our data from various different sources. The language model is trained on the English side of entire data (1.65M sentences, which is 39.3M words.)

Source	解决 13 亿 人 的 问题 , 不能 靠 别人 , 只能 靠 自己 .
Reference	to solve the problem of 1.3 billion people , we can only rely on ourselves and nobody else .
Baseline	cannot rely on others , can only resolve the problem of 13 billion people , on their own .
Method 2	to resolve the issue of 1.3 billion people , they can't rely on others , and it can only rely on themselves .

Source	在 新世纪 新 形势 下 , 亚洲 的 发展 面临 着 新 的 机遇 .
Reference	in the new situation of the millennium , the development of asia is facing new opportunities .
Baseline	facing new opportunities in the new situation in the new century , the development of asia .
Method 2	under the new situation in the new century , the development of asia are facing a new opportunity .

Source	他 说 , 阿盟 是 同 美国 讨论 中东 地区 进行 民主改革 的 最佳 伙伴 .
Reference	he said the arab league is the best partner to discuss with the united states about carrying out democratic reforms in the middle east .
Baseline	arab league is the best with democratic reform in the middle east region in the discussion of the united states , he said .
Method 2	arab league is the best partner to discuss the middle east region democratic reform with the united states , he said .

Table 2: Comparison of example translations from the baseline method and our Method 2.

The best BLEU score on the test set is 25.92, which is from the experiments of Method 2. Method 1 system seems to behave slightly worse than the baseline and Method 2. The reason for this behavior is that the rules that were extracted from the semantic role labeled corpus could have isolated semantic roles in them which would not necessarily get connected to the right predicate or argument during decoding. In other words, it is possible for a rule to only contain one or some of the semantic arguments of a predicate, and not even include the predicate itself, and therefore there is no guarantee that the predicate will be translated with the right arguments and in the right order. The difference between the BLEU scores of the best Method 2 results and the baseline is 0.92. This improvement is statistically significant ($p = 0.032$) and it shows that incorporating semantic roles in machine translation is an effective approach.

Table 2 compares some translations from the baseline decoder and our Method 2. The first line of each example is the Chinese source sentence, and the second line is one of the reference translations. The last two lines compare the baseline and Method 2. These examples show how our Method 2 can outperform the baseline method, by translating complete semantic structures, and generating the semantic roles in the correct order in the target language. In the first example, the predicate *rely on* for the argument *themselves* was not translated by the baseline decoder, but it was correctly translated by Method 2. The second example is a case where the baseline method generated the arguments in the wrong order (in the case of *facing* and *development*), but the translation by Method 2 has the correct order. In the last example we see that the arguments of the predicate *discuss* have the wrong order in the baseline translation,

	BLEU Score	
	dev	test
Baseline	26.01	25.00
Method 1	26.12	24.84
Method 2	26.5	**25.92**

Table 3: BLEU scores on the test and development sets, of 8 experiments with random initial feature weights.

but Method 2 generated the correct oder.

4 Conclusion

We proposed two methods for incorporating semantic role labels in a string-to-tree machine translation system, by learning translation rules that are semantically enriched. In one approach, the system learned the translation rules by using a semantic role labeled corpus and augmenting the set of nonterminals used in the rules, and in the second approach, in addition to the regular SCFG rules, the system learned semantic roles which contained the complete semantic structure of a predicate, and added a feature to distinguish those rules.

The first approach did not perform any better than the baseline, which we explained as being due to having rules with only partial semantic structures and not having a way to guarantee that those rules will be used with each other in the right way. The second approach significantly outperformed the baseline of our experiments, which shows that complete predicate-argument structures can improve the quality of machine translation.

References

Kathryn Baker, Michael Bloodgood, Chris Callison-Burch, Bonnie J. Dorr, Nathaniel W. Filardo, Lori Levin, Scott Miller, and Christine Piatko. 2010. Semantically-informed machine translation: A tree-grafting approach. In *Proceedings of The Ninth Biennial Conference of the Association for Machine Translation in the Americas*, Denver, Colorado.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Tagyoung Chung, Licheng Fang, and Daniel Gildea. 2011. Issues concerning decoding with synchronous context-free grammar. In *Proceedings of the ACL 2011 Conference Short Papers*, Portland, Oregon. Association for Computational Linguistics.

Hagen Fürstenau and Mirella Lapata. 2012. Semi-supervised semantic role labeling via structural alignment. *Computational Linguistics*, 38(1):135–171.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In *Proceedings of NAACL-04*, pages 273–280, Boston.

Daniel Gildea and Daniel Jurafsky. 2000. Automatic labeling of semantic roles. In *Proceedings of ACL-00*, pages 512–520, Hong Kong, October.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of NAACL-03*, pages 48–54, Edmonton, Alberta.

Beth Levin. 1993. *English Verb Classes And Alternations: A Preliminary Investigation.* University of Chicago Press, Chicago.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *COLING-10*, Beijing.

Franz Josef Och. 2003. Minimum error rate training for statistical machine translation. In *Proceedings of ACL-03*, pages 160–167.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. *Semantic Role Labeling.* Synthesis Lectures on Human Language Technology Series. Morgan and Claypool.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of ACL-02*, pages 311–318.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440, Sydney, Australia, July. Association for Computational Linguistics.

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James Martin, and Dan Jurafsky. 2004. Shallow semantic parsing using support vector machines. In *Proceedings of NAACL-04*.

V. Srikumar and D. Roth. 2011. A joint model for extended semantic role labeling. In *EMNLP*, Edinburgh, Scotland.

Kristina Toutanova, Aria Haghighi, and Christopher Manning. 2005. Joint learning improves semantic role labeling. In *Proceedings of ACL-05*, pages 589–596.

Wei Wang, Jonathan May, Kevin Knight, and Daniel Marcu. 2010. Re-structuring, re-labeling, and re-aligning for syntax-based machine translation. *Computational Linguistics*, 36:247–277, June.

Dekai Wu and Pascale Fung. 2009. Semantic roles for smt: A hybrid two-pass model. In *Proceedings of the HLT-NAACL 2009: Short Papers*, Boulder, Colorado.

Xianchao Wu, Takuya Matsuzaki, and Jun'ichi Tsujii. 2010. Fine-grained tree-to-string translation rule extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, Stroudsburg, PA, USA. Association for Computational Linguistics.

Deyi Xiong, Min Zhang, and Haizhou Li. 2012. Modeling the translation of predicate-argument structure for smt. In *ACL (1)*, pages 902–911.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of EMNLP*.