Injecting a Semantic Objective Function into Early Stage Learning of Spoken Language Translation

Meriem Beloucif and Dekai Wu
Human Language Technology Center
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
Email: {mbeloucif | dekai@cs.ust.hk}

Abstract—We describe a new approach for semantically training spoken language translation systems, in which we inject a crosslingual semantic frame based objective function directly into inversion transduction grammar (ITG) induction. This represents an ambitious jump from recent work on improving translation adequacy by using a semantic frame based objective function to drive the tuning of loglinear mixture weights in the final stage of statistical machine translation training. In contrast, our new approach propagates a semantic frame based objective function back into much earlier stages of the pipeline, during the actual learning of the translation model, biasing learning toward semantically more accurate alignments. This approach is motivated by the fact that ITG alignments have empirically been shown to fully cover crosslingual semantic frame alternations, even though they rule out an overwhelming majority of the space of possible alignments. We show that directly driving ITG induction with a crosslingual semantic based objective function not only helps to further sharpen the ITG constraints, but still avoids excising relevant portions of the search space, leading to better performance than either conventional ITG or GIZA++ based approaches.

I. Introduction

In this paper we describe a new approach that uses a crosslingual semantic based objective function at a very early stage of training spoken language translation systems. Recent research has shown that including a semantic based objective function in the training pipeline by tuning against semantic based metrics, MEANT [1]), improves the translation adequacy [2]–[5]. We show that integrating a semantic based objective function much earlier in the training produces a more semantically correct alignment. Our approach is motivated by the success of our recently developed crosslingual evaluation metric, XMEANT [6]. We employ an XMEANT based crosslingual semantic frame alignment method for constraining inversion transduction grammars (ITGs). We show that this way of inducing ITGs helps to learn more semantically valid alignments compared to both conventional ITGs and the traditional GIZA++ alignments, leading to better translations. Our approach is motivated by the fact that XMEANT has been shown to correlate better with human adequacy judgement than most of the commonly used metrics [6]. Furthermore, ITG alignments have previously been empirically shown to almost fully cover crosslingual semantic frame alternations, even though they rule out the majority of incorrect alignments [7]. We show that using XMEANT-like semantic frame matching for inducing ITGs not only helps to further narrow down inversion transduction grammar constraints, but also avoids losing relevant portions of the search space, leading to more semantically driven word alignments. We also show that a semantic based learning can also help to improve the translation quality for low resource languages in comparison to existing learning methods by deliberately training our approach using a relatively small dataset. We adopt DARPA's approach in the LORELEI dry run evaluation, simulating low resource conditions in a Chinese-English translation learning task (despite the fact that Chinese is not a low resource language) by deliberately restricting the parallel training data to a small dataset, namely the International Workshop on Spoken Language Translation (IWSLT07) spoken language translation corpus, and show that our method outperforms the traditional alignment methods for spoken data.

II. Related work

A. The MEANT family of metrics

Our method is fully compatible with the principle adopted by the MEANT family of metrics, in which a good output translation is deemed to be one that preserves the core semantic frames of the input sentence as captured by the basic event structure who did what to whom, for whom, when, where, how and why [8]. Recent work has shown that the semantic frame based metric MEANT correlates better with human adequacy judgment than most common surface form based evaluation metrics [1], [9], [10] such as BLEU [11], NIST [12], METEOR [13], CDER [14], WER [15], and TER [16].

MEANT is a weighted f-score over the matched semantic role labels of automatically aligned semantic frames and role fillers [1], [9], [10]. It evaluates the degree of goodness of the MT output sentence against the provided reference translations, and produces a score that measures the degree of similarity between their semantic frame structures. Our new approach is encouraged by the fact that many previous studies have empirically shown that integrating semantic role labeling into the training pipeline by tuning against MEANT improves the translation adequacy [2]–[5].

B. XMEANT: crosslingual MEANT

The crosslingual XMEANT metric [6] has been shown to correlate even better with human adequacy judgments than
MEANT. Unlike MEANT, which needs the expensive man made references for the MT evaluation, XMEANT uses the foreign input to evaluate the MT translation output. MEANT measures lexical similarity using a monolingual context vector model, whereas XMEANT substitutes simple crosslingual lexical translation probabilities. Figure 2 describes the XMEANT algorithm. XMEANT uses MEANT’s f-score based method for aggregating lexical translation probabilities within the semantic role filler phrases. Each token of the role fillers in the output/input string is aligned to the token of the role fillers in the input/output string that has the maximum lexical translation probability. The crosslingual phrasal similarities are computed as shown in figure 1.

Our approach uses the XMEANT method of matching semantic predicates and role labels between the input and the output, and uses this crucial information for inducing inversion transduction grammars. In this paper we show that by using this semantic objective function at an early stage of training the statistical machine translation (SMT) system, not only are we able to learn more semantic correlations between the two languages, but also that this holds even under low resource conditions limited to small amounts of parallel data, as in the DARPA LORELEI program.

C. Alignment

Word alignment is considered to be an important step in training MT systems, since it helps to learn the correlations between the input and the output languages. Unfortunately, conventional alignments are generally based on training IBM models [17], which are known to produce weak word alignment since they allow unstructured movement of words. Then take the intersection of alignments in both directions to produce the final alignment. A hidden Markov model (HMM) based alignment was proposed by Vogel et al. [18], but similarly to IBM models, the objective function uses surface based alignment rather than a more structure based alignment. No constraints are used while training, allowing many random word-to-word permutations. Such an alignment generally hurts translation accuracy and adequacy.

For producing word alignments via unsupervised training of inversion transduction grammars [19], a method with improved efficiency has been developed in work starting with Saers et al. [20]. This method tackles the issue that exhaustive biparsing and training using ITGs requires $O(n^6)$ time which, though feasible, is slow; instead, an improved method runs in $O(n^3)$ time [21].

In this work, we use BITGs or bracketing transduction grammars [20], which only use one single nonterminal category and surprisingly achieve good results by outperforming the conventional GIZA++ alignments [22]. It has been shown that ITG constraints allow higher flexibility in word ordering for longer sentences than the conventional IBM model, and that applying ITG constraints for word alignment leads to learning a significantly better alignment than the constraints used in conventional IBM models for both German-English and French-English language pairs [23]. In a version of ITG proposed by Zhang and Gildea [24], rule probabilities are lexicalized throughout the biparse tree for efficient training, which helped align sentences up to 15 words. Some of the previous work on word alignment used morphological and syntactic features [25]. Some log linear models have been proposed to incorporate these features [26]. The problem with these approaches is that they require language specific knowledge and that they always work better on more morphologically rich languages.

Few studies that approximately integrate semantic knowledge in computing word alignment have been proposed [27], [28]. However, the former needs to have a prior word alignment learned on lexical words. The authors in the latter model proposed a semantic oriented word alignment. However, they need to extract word similarity from the monolingual data first then produce alignment using word similarities.

III. Using XMEANT similarity for ITGs induction

The model we propose in this paper injects an XMEANT semantic frame based objective function into early stage SMT training, thereby biasing the bracketing inversion transduction grammar (BITG) towards preferring bilingual constituents that best fits XMEANT’s crosslingual semantic frames. Owing to the structural differences between the monolingual semantic parsers and the bilingual BITG parses, XMEANT
rewards/penalizes BITG biconstituents fitting or violating the crosslingually aligned semantic frames.

The semantic roles and their fillers in a sentence sometimes span across multiple syntactic units, or in technical terms: the semantic trees are not necessarily consistent with the syntactic trees. Since BITG trees are defined to be projective, applying even a single monolingual semantic parse would rule out all possible BITG trees, and all possible alignments for that sentence pair. As the lexical relation is what defines the word alignment, which is what we are interested in, XMEANT penalizes any constraints that violate XMEANT's semantic frame alignment. In practice, the automatic semantic shallow parses are fairly noisy, which, in engineering, is a reason to soften them, but even with perfect semantic parsers, additional constraints are theoretically necessary.

A penalty is paid whenever the BITG biparser wants to introduce a biconstituent that does not fit into XMEANT's alignment, or in other words, crosses a semantic constituent (the string of a predicate or one of its role fillers span). In this way, a reward is paid for biconstituents completely covering a semantic constituent or by biconstituents that are completely covered by semantic constituents. To allow for some degree of freedom, we allow for two penalty levels, one for crossing an input language semantic constituent, and one for crossing an output language semantic constituent. The same applies for the reward constituent. These hyperparameters need to be set manually for now, we are studying a smarter way to set these hyperparameters.

IV. Experimental Setup

A. Word alignment

We compare the performance of our proposed XMEANT-driven alignments to the conventional ITG alignment and to the traditional GIZA++ baseline with grow-diag-final-and to harmonize both alignment directions. We also perform a grid search over the hyper parameters in our proposed model to find the optimal settings.

Our ITG baseline is a token-based BITG system. We initialize it with uniform structural probabilities, setting aside half of the probability mass for lexical rules. This probability mass is distributed among the lexical rules according to co-occurrence counts from the training data, assuming each sentence to contain one empty token to account for singletons. These initial probabilities are refined with 10 iterations of expectation maximization where the expectation step is calculated using beam pruned parsing [22] with a beam width of 100. On the last iteration, we extract the alignments imposed by the Viterbi parses as the word alignments outputted by the system.

Compared to the ITG baseline discussed above, our new model rewards any biconstituent that falls into XMEANT semantic frame alignment, as discussed in Section 3. The shallow semantic parses of the training data were produced using ASSERT [8] and C-ASSERT [29] for English and Chinese respectively. The hyperparameters were only used during training to set the probabilities of the grammar, not when extracting the Viterbi parses and the corresponding word alignments.

B. SMT pipeline

In our experiments, we purposely use a relatively small corpus to simulate low resource language scenario. We show that including a semantics based objective function during the actual learning of the SMT model helps better learning bilingual correlations, without relying on heavy memorization from expensive huge parallel corpora. Although Chinese is not a low resource language, we adopted the DARPA LORELEI program's approach in its dry run evaluation, by purposely simulating low resource conditions, in the present case by using a relatively small corpus (IWSLT07). The training set contains 39,953 sentences. The training set, development set, and test set were the same for all systems in order to keep the experiments comparable.

We tested the different alignments described above by using the standard MOSES toolkit [30], and a 6-gram language model learned with the SRI language model toolkit [31] to train our model. We tested our approach with both MOSES hierarchical and MOSES phrase-based. For tuning, we used ZMERT [32] the standard implementation of minimum error rate training, or MERT [33].

V. Results

We compared the performance of semantic based BITG alignment to the GIZA++ baseline and the conventional BITG for both MOSES hierarchical and MOSES phrase based. We evaluated our MT output using the semantic metric MEANT [1] and also surface based metrics such as BLEU [11], METEOR [13], CEDER [14], WER [15], and TER [16]. We observe that both ITG based systems give a comparable result which is still very high in comparison to GIZA++ alignment in term of edit distance metrics and MEANT score. Tables I and II show the interesting improvement in terms of BLEU and MEANT scores for our proposed XMEANT-driven aligned system in comparison to conventional BITG alignment and GIZA++ alignment for both Moses baselines. Both BLEU and MEANT scores for our new proposed alignment are considerably higher than the BLEU and MEANT scores for the conventional BITG and the traditional GIZA++ based systems. We also note that the MEANT score for ITG with semantic constraints is slightly better than the conventional ITG model. We believe that a better shallow semantic parser would yield a better system. Our results show that we should be more focused on including semantic information while training SMT system rather than just tuning against a semantic objective function.

On the other hand, Figure 4 shows an interesting example extracted from our translated data and compared to the translations obtained by other systems. We see from the example that ITG based models give a more accurate output compared to GIZA++ based alignment. Example 1 shows an interesting example in which the XMEANT-driven system learns a more accurate translation of the input sentence,
I am going to stay at the dormitory of Harvard University.

我 会 住 在 哈 佛 大 学 的 宿 舍。

Fig. 3. An alignment of bi-sentences produced by both GIZA++ (left) and ITG based alignments (right) at the top of the picture, and the XMEANT-driven constrained ITG alignment at the bottom.

Example 1
Input: 补牙 的 填充 物 脱落 了。
Ref: a filling has come out.
GIZA++: the tooth has.
ITG: a filling has come off behind.
XMEANT_based: lost a filling behind.

Example 2
Input: 食堂 在 哪里 ？
Ref: where's the dining room?
GIZA++: refectory then where?
ITG: the refectory where?
XMEANT_based: where's the refectory?

Example 3
Input: 能 告诉 我 登记 时间 吗 ？
Ref: could you tell me the boarding time, please?
GIZA++: can I check in?
ITG: can you tell me the check-in time?
XMEANT_based: can you tell me the business hours?

Fig. 4. Interesting examples comparing the output of the three compared systems.
whereas the GIZA++ fails completely to capture the basic semantics of the input. The ITG system on the other hand, correctly gets the global meaning of the input but fails to use the right wording (has come off). Example 2 shows an example where learning the right semantic structure can not only produce better adequacy, but also leads to better fluency for low resource languages. We emphasize here, that both GIZA++ and ITG models fail to capture the right translation due to insufficient training data. The semantic frame based objective function that we used shows that by capturing the right structure while learning the alignment, we can produce better translations even when using a very small data set. Example 3 is also interesting in the sense that, having no context, both ITG and XMEANT output can be considered as valid translations. This shows again, that semantic based heuristics are needed for more disambiguation, on the other hand, GIZA++ based alignment fails to completely capture any meaning once again.

Figure 3 represents the alignment obtained after running GIZA++, the ITG based system, and our new system baseline respectively. We observe that both GIZA++ and ITG alignments fail to align different crucial part of the parallel sentences. The XMEANT-driven alignment gives a very good alignment based on the semantic structure of both semantic parsers. We see that it only fails while trying to align the to ®, which can be explained by the fact that, from either English-to-Chinese or Chinese-to-English, the word the or the character ¤ will be translated to NULL. There are cases where ® gets translated to other similar non-function-words such as ‘s or quotation marks, but we can consider these to detract relatively little from the general understandability of the translation.

VI. Conclusion

In this paper we have presented an approach to semantically drive the learning of spoken language translation models, by using the XMEANT objective function to drive early stage ITG induction. We show that including a semantic based objective function at an early stage of the SMT pipeline helps to improve both the fluency and the adequacy of the machine translation. We have also demonstrated that using XMEANT constraints in ITG alignment produces a more semantically correct alignment and thus yields interesting improvements compared to conventional ITG alignment and to the traditional GIZA++ alignment.

Finally, we also tested the performance of our model against MOSES hierarchical and MOSES phrase based translation baselines. We observed that systems using our semantically based approach for word alignment are comparable to BITG alignment systems in terms of edit distance metrics like TER, WER, PER and CDER, and that they both highly outperform the GIZA++ alignment based system results for Chinese to English translations.

Acknowledgments

This material is based upon work supported in part by the Defense Advanced Research Projects Agency (DARPA) under LORELEI contract HR0011-15-C-0114, BOLT contracts HR0011-12-C-0014 and HR0011-12-C-0016, and GALE contracts HR0011-06-C-0022 and HR0011-06-C-0023; by the European Union under the Horizon 2020 grant agreement 645452 (QT21) and FP7 grant agreement 287658; and by the Hong Kong Research Grants Council (RGC) research grants GRF16210714, GRF16214315, GRF620811 and GRF621008. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA, the EU, or RGC.

References


TABLE I

<table>
<thead>
<tr>
<th>System</th>
<th>MEANT</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
<th>CDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giza alignment</td>
<td>49.94</td>
<td>23.02</td>
<td>4.14</td>
<td>59.95</td>
<td>60.52</td>
<td>55.58</td>
<td>59.14</td>
</tr>
<tr>
<td>ITG alignment</td>
<td>50.57</td>
<td>21.82</td>
<td>4.32</td>
<td>57.86</td>
<td>58.68</td>
<td>53.90</td>
<td>57.38</td>
</tr>
<tr>
<td>XMEANT-driven</td>
<td>50.92</td>
<td>24.70</td>
<td>4.27</td>
<td>58.44</td>
<td>59.01</td>
<td>53.85</td>
<td>57.58</td>
</tr>
</tbody>
</table>

TABLE II
Translation quality of the three alignment methods used in Chinese-English MT systems using IWSLT 2007, trained using Moses phrase based.

<table>
<thead>
<tr>
<th>System</th>
<th>MEANT</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
<th>CDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giza alignment</td>
<td>47.65</td>
<td>18.59</td>
<td>3.70</td>
<td>63.04</td>
<td>63.83</td>
<td>57.37</td>
<td>62.02</td>
</tr>
<tr>
<td>ITG alignment</td>
<td>48.36</td>
<td>18.44</td>
<td>4.02</td>
<td>61.14</td>
<td>62.63</td>
<td>54.96</td>
<td>60.54</td>
</tr>
<tr>
<td>XMEANT-driven</td>
<td>48.56</td>
<td>20.35</td>
<td>4.02</td>
<td>61.17</td>
<td>62.77</td>
<td>55.42</td>
<td>60.46</td>
</tr>
</tbody>
</table>


