Pre-ordering of phrase-based machine translation input in translation workflow

Alexandru Ceausu
Euroscript Luxembourg S.à r.l.
Bertrange, Luxembourg
alexandru.ceausu@euroscript.lu

Sabine Hunsicker
euroscript Deutschland GmbH
Berlin, Germany
sabine.hunsicker@euroscript.de

Abstract

Word reordering is a difficult task for decoders when the languages involved have a significant difference in syntax. Phrase-based statistical machine translation (PBSMT), preferred in commercial settings due to its maturity, is particularly prone to errors in long range reordering. Source sentence pre-ordering, as a pre-processing step before PBSMT, proved to be an efficient solution that can be achieved using limited resources. We propose a dependency-based pre-ordering model with parameters optimized using a reordering score to pre-order the source sentence. The source sentence is then translated using an existing phrase-based system. The proposed solution is very simple to implement. It uses a hierarchical phrase-based statistical machine translation system (HPBSMT) for pre-ordering, combined with a PBSMT system for the actual translation. We show that the system can provide alternate translations of less post-editing effort in a translation workflow with German as the source language.

Keywords: pre-ordering, statistical machine translation, reordering scores

1. Introduction

Machine translation (MT) was adopted as a productivity tool in most translation workflows. Translation memory (TM) enrichment is a common usage scenario employed by large language service providers. The delivery of machine translation output as an alternative translation suggestion in a familiar CAT (computer-assisted tool) environment is less intrusive for translators than post-editing a pre-translated document. In this setting, different candidates from machine translation can be provided in the same working environment.

One of the most obvious issue that influence post-editing effort is due to the MT output inconsistency in syntax.

Word reordering is a difficult task for decoders when the languages involved have a significant difference in syntax. Phrase-based statistical machine translation (PBSMT), preferred in commercial settings due to its maturity, is particularly prone to errors in long range reordering. Source sentence pre-ordering, as a pre-processing step before PBSMT, proved to be an efficient solution that can be achieved using limited resources.

Post-editing a machine translation (MT) output, where long-range reordering phenomena was not correctly solved by MT, requires a considerable effort from the translator. The accuracy of long-range in MT plays an important role in the decision of the translator if the segment is fit for post-editing or if it is better to be translated from scratch.

In order to address these, we propose a dependency-based pre-ordering model with parameters optimized using a reordering score to pre-order the source sentence. We tested the solution for the German–English language pair. The system provides alternate translations in a translation workflow in which a phrase-based decoder is available. The pre-requisites for this task are: (i) no retraining of the phrase-based model on pre-ordered data, (ii) only long range re-ordering is allowed (assuming that the phrase-based system has an adequate handling of local reordering). This solution has a practical approach: the sub-tree permutation is handled by the Moses chart decoder using synchronous context-free grammar rules, and it is tuned using minimum error rate training (MERT) with a reordering score (Kendall Tau distance or Kendall Reordering Score - KRS). The alignments are used to pre-order the input for the phrase-based decoder.

2. Background

If we take the English language as the reference, there are several other languages for which the PBSMT system has significant problems when dealing with long range reordering. Among European languages, German is one for which this problem has to be addressed.

The language pair under investigation in this experiment is German–English. For this language pair, the most significant difference relates to verb positioning. In German, modal verbs and verbs in subordinate clauses have different positions than in English, as in this example:

German:
Fürsprecher der Legalisierungsbewegung *hoffen*$_1$, dass das Justizministerium *nachgeben*$_2$ wird.

English:
Advocates for legalized marijuana are hoping$_1$ the Justice Department *yields*$_2$.

We are also focusing on other phenomena of syntactic order divergence occurring in German to English, including subject movement, verb article contractions and negations (Collins et all, 2005):

German:
Die Kinder können *andere*$_1$ bekommen.
English:
Other$_1$ people$_1$ can have children.

3. Related work

Reordering for language pairs with considerably different syntax is an active topic of research in statistical machine translation. For these language pairs, MT has to solve the difficult problem of long-range reordering, a problem more difficult to solve than local reordering. We can differentiate the solutions addressing the long-range reordering problem by the location where they are applied. There are approaches that address this problem directly in the decoder; some others solve it through pre-processing and/or post-processing steps. Another criterion would be to distinguish between rule-based and statistical solutions.

There are several approaches which we can use as a baseline for pre-ordering.

Particularly interesting for our approach are the experiments regarding German to English comparison of pre-ordering methods of Navrátil et al, 2012; the experiments on automatic source-language syntactic pre-processing for Arabic-English described in Habash, 2007 and the experiments in parsing for Japanese-English syntactic-reordering (Katz-Brown et al, 2008, 2011).

A comprehensive enumeration of long range reordering techniques for the German–English language pair is available in Bisazza and Frederiko, 2013. Especially interesting for the German–English language pair is the reordering score proposed in this paper: the verb-specific Kendall Reordering Score. This score is a version of the Kendall Reordering score that focuses on verb long-range reordering.

Important for optimizing the parameters for the pre-ordering system are the works of Alexandra Birch and Miles Osborne (2010) regarding reordering metrics for machine translation.

4. Data preparation

We used all freely available data (see Table 1) to build the German–English system. The data is made available from several research projects, the EUROPARL corpus (Koehn, 2005), the DG Translation Memory (Steinberger, 2012), the ECB and EMEA corpora from the OPUS parallel corpus (Tiedmann, 2012), etc.

The corpus was pre-processed using the tools provided by the Moses project [1]. Among these pre-processing steps we can mention: parallel corpus cleaning, tokenisation, true-casing, compound-splitting, etc.

For parsing the German data we used the non-projective dependency parser (Bohnet and Nivre, 2012) from the *mate-tools* [2].

As an additional step after word alignment, we discarded the one-to-many alignments that did not have an equivalent in the dependency relations. We had this additional pre-processing step in order to make rule extraction less ambiguous. This can be achieved if both source and target are dependency parsed.

Corpus	Sentence pairs
DG-Translation Memory[3]	4033963
EUROPARL-v7[4]	1920209
news-commentary-v7[5]	158840
multiun doc[6]	162981
EMEA[7]	1108752
ECB[8]	113174
Total	**7497919**

Table 1. Available data for the German-English system

5. Rule extraction

For reordering rule extraction we assume that we can convert word alignments into permutations. The alignments map source words to target positions. As alignments can be ambiguous, allowing null, one-to-many and many-to-one relations, we need a simple mechanism to convert them to permutations. We are using the same simplification algorithm described in (Birch and Osborne 2010). Source words aligned to null receive the incremented position of the previous aligned source word. In many-to-one alignments, source words receive monotone target positions. One-to-many alignments map the source word to the first target position.

Using dependency parses for German and the word alignments, we can automatically generate sub-tree reordering rules. Each rule is labelled with the name of the dependency relation. The head of the relation is represented using its part-of-speech and the dependent nodes are represented using the label of their relation. The terminals are represented with their part-of-speech label. We choose to flatten the non-projective dependency links and assign them to the start rule. Here is an example on how a participle (VAPP), a predicate (PD) and a modifier (MO) relation can be reordered in a clausal object relation (OC).

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

[2] http://code.google.com/p/mate-tools/
[3] http://ipsc.jrc.ec.europa.eu/?id=197
[4] http://www.statmt.org/europarl/
[5] http://www.statmt.org/wmt13/translation-task.html
[6] http://www.euromatrixplus.net/multi-un/
[7] http://opus.lingfil.uu.se/EMEA.php
[8] http://opus.lingfil.uu.se/ECB.php

```
OC -> (PD_1 VAPP MO_1, MO_1 VAPP PD_1)
OC -> (PD_1 VAPP MO_1, PD_1 MO_1 VAPP)
OC -> (PD_1 VAPP MO_1, PD_1 VAPP MO_1)
OC -> (PD_1 VAPP MO_1, VAPP PD_1 MO_1)
```

Each rule has several scores associated to it, which receive a weight during tuning. One of them is the maximum likelihood estimate (MLE) from the relative frequencies of the rules in the training corpus. Other scores account for rule length, for the relative and absolute distance in sentence between head and dependent, etc.

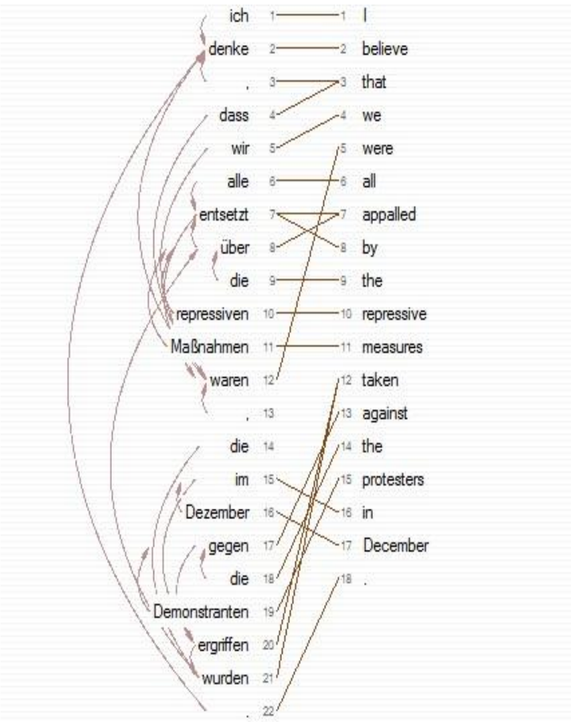

Figure 1. Dependency annotated German side of a sentence pair

In Figure 1 we show an example of directed dependency annotation for the German side of the corpus. The example shows clearly the movement of the verb in the sub-clauses by the crossed lines.

6. Feature extraction

Feature extraction, in our case, is the process of assigning additional scores to the existing rules. These scores are designed to assist the parameter optimization step to produce a HPBSMT system better at predicting long-range reordering. Among these, we can mention: scores based on the Kendall Reordering Score, scores based on a deviation from the median rule length, scores based on the part-of-speech (POS) involved, etc.

The Rule KRS assesses how similar is the input -output alignment as to the input-reference alignment. As the rule has the target terminals and non-terminals aligned to source ones, the score can be applied to the source and target side of the rule. In Birch and Osborne (2010), the Kendall Reordering Score is described as:

$$d_\tau(\pi, \sigma) = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij}}{Z}$$

$$\text{where } z_{ij} = \begin{cases} 1 \; if \; \pi(i) < \pi(j) and \; \sigma(i) > \sigma(j) \\ 0 \; otherwise \end{cases}$$

$$Z = \frac{(n^2 - n)}{2}$$

Modified Rule KRS is a modified KRS score that only take into consideration the reordering of the difficult POS tags.

Rule length deviation accounts for how much the rule differs in length from other rules. In our experiments, this score is tight to the dependency relation governor.

The Governor POS score measures how likely it is for a particular governor type to take part in a long range reordering.

7. Parameter optimisation

Parameter optimisation is done using MERT (Och and Ney, 2002). MERT can be used with several metrics, the most employed metric being the BLEU score (Papineni et al, 2001).

As shown by Birch and Osborne (2010), the BLEU score is not appropriate to assess the translation quality when long range reordering is involved. Instead, we are using the Kendall Reordering Score as proposed by Alexandra Birch and Miles Osborne (2010). In the former mentioned paper, the Kendall Reordering Score (KRS) is interpolated with the BLEU score for better correlation with human judgments. KRS penalizes distortion between source and target alignments.

The Kendall Reordering Score, as well as a derivation of it that advantages long reordering jumps for verbs, was successfully used to directly reorder the words in a German–English experiment (Bisazza and Frederico, 2013).

The corpus used for parameter optimisation and testing was produced from post-edited translation segments of several documents. The documents were first translated using machine translation and then post-edited by professional translators. We only selected sentences longer than 25 tokens in which we can observe long-range reordering patterns. We cumulated 4000 sentences, out of which we used 2000 for testing and 2000 for parameter optimisation.

8. Maximum likelihood approach

The baseline approach uses only the score that yields the relative frequency of a rule, choosing the rule with the highest frequency. The baseline pre-ordering does not need a parameter optimisation step. There are no other scores as the source lexicon has a perfect match to the target lexicon. The hierarchical model we use for reordering does not need a language model.

Parameter tuning is used to assign weights to all the scores that influence the quality of translation, including scores for rule length, the relative and absolute distance in sentence between head and dependent, if the head is the subject, if the head is the verb, if it is a clause with negation, etc.

The same PBSMT system is used for all three runs. For the Baseline MLE, the PBSMT system uses as input the pre-ordered target sentence of the baseline MLE. For the Tuned MLE setting, the input of the PBSMT engine is the pre-ordered target sentence of the Tuned MLE system.

Table 2 presents the results of the experiment where we compared the baseline MLE pre-ordering as opposed to the tuned approach. As it can be seen, the BLEU does not show a considerable change between the baseline and the tuned MLE systems. As opposed to that, the edit distance score, computed in the first phase of the experiment from the human post-edits of the PBSMT output, shows a statistically significant increase for the tuned MLE system.

	Edit distance	**BLEU**
PBSMT	78.33	46.23
Baseline MLE pre-ordering	79.87	47.05
Tuned MLE pre-ordering	82.07	48.52

Table 2. Edit distance and BLEU scores for PBSMT, the baseline and the tuned MLE system

9. Conclusions

During our experiments with pre-ordering we found that long-range reordering using the proposed pre-ordering technique reduces the edit distance score, and, subsequently, the post-editing effort.

The same algorithm can be applied to other language pairs as well. We are currently extending the experiment to Dutch and Danish.

The experiment also produced labelled data of long range reordered sentences. We are extending this corpus with annotated data from other Germanic languages, like Dutch and Danish.

References

Birch A., Osborne M., 2011. Reordering Metrics for MT. ACL, Portland, Oregon, USA

Birch A., Osborne M., and Koehn P.. 2008. Predicting success in machine translation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 745– 754, Stroudsburg, PA, USA.

Bisazza A. and Federico M., 2013. Efficient solutions for word reordering in German-English phrase-based statistical machine translation, in Proceedings of the Eighth Workshop on Statistical Machine Translation, pages 440–451, Sofia, Bulgaria, August 8-9, 2013

Bohnet B., and Nivre J., 2012 *A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing.* EMNLP-CoNLL, pages 1455-1465

Collins M., Koehn P., and Ucerova I. 2005. Clause restructuring for statistical machine translation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pages 531–540, Ann Arbor, Michigan, June.

Habash, N., 2007. Syntactic preprocessing for statistical machine translation, Proceedings of the 11th MT Summit

Katz-Brown J., Collins M., 2008. Syntactic reordering in preprocessing for Japanese-to-English translation: MIT system description for NTCIR-7 patent translation task, NTCIR 2008

Katz-Brown J., Collins M., 2008. Syntactic reordering in preprocessing for Japanese-to-English translation: MIT system description for NTCIR-7 patent translation task, NTCIR 2008

Katz-Brown J., Petrov S., McDonald R., Och F., Talbot D., Ichikawa H., Seno M., 2011. Training a Parser for Machine Translation Reordering, EMNLP 2011

Koehn P., Europarl: A Parallel Corpus for Statistical Machine Translation, MT Summit 2005

Navrátil, J., Visweswariah K., Ramanathan A., 2012. A Comparison of Syntactic Reordering Methods for English-German Machine Translation, Proceedings of COLING 2012: Technical Papers, pp. 2043–2058

Och F. and Ney H., 2002. Discriminative training and maximum entropy models for statistical machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pages 295–302, Philadelhpia, PA.

Papineni K., Roukos S., Ward T., and Zhu W.-J.. 2001. Bleu: a method for automatic evaluation of machine translation. Research Report RC22176, IBM Research Division, Thomas J. Watson Research Center.

Steinberger R., Eisele A., Klocek S., Pilos S. & Schlüter P. (2012). DGT-TM: A freely Available Translation Memory in 22 Languages. Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012), Istanbul, 21-27 May 2012

Tiedemann J., 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)