The Karlsruhe Institute for Technology Translation System for the ACL-WMT 2010

Jan Niehues, Teresa Herrmann and Alex Waibel
Karlsruhe Institute for Technology
Karlsruhe, Germany
{jan.niehues,teresa.herrmann,alex.waibel}@kit.edu

Abstract

1 Introduction

In this paper we describe the systems that we built for our participation in the Shared Translation Task of the ACL 2010 Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR. Our translations are generated using a state-of-the-art phrase-based translation system and applying different extensions and modifications including a POS-based reordering model, a factored translation model using POS and stem information as well as bilingual language models based on the same type of linguistic word-level information.

Depending on the source and target languages, the proposed models differ in their benefit for the translation task and also expose different correlative effects. The following sections introduce the characteristics of the baseline system and the supplementary models. In Section 5 we present the performance of the system variants applying the different models and chose the systems used for creating the submissions for the English-German and German-English translation task. Section 6 draws conclusions and suggests directions for future work.

2 Baseline System

The baseline systems for the translation directions German-English and English-German are both developed using discriminative word alignment (Niehues and Vogel, 2008), the Moses Toolkit (Koehn et al., 2007) for extracting phrase pairs and generating the phrase table from the discriminative word alignments. Translations are performed by the STTK Decoder (Vogel, 2003) and all systems are optimized towards BLEU using Minimum Error Rate Training (Venugopal et al., 2005).

2.1 Training, Development and Test Data

We used the data provided for the WMT10 for training, optimizing and testing our systems: Our training corpus consists of Europarl and News-Commentary data, for optimization we use newstest2008 as development set and as test set we use newstest2009.

The baseline language models are trained on the target language part of the Europarl and News-Commentary corpora. Additional, bigger language models were trained on the monolingual corpora: For German as target language the News corpus was used while for English as target language the even bigger Gigaword corpus was available.

2.2 Preprocessing

The training data was preprocessed before used for training. In this step different normalizations like mapping different types of quotes were done. In the end the first word of every sentence was smart-cased.

For the German text additional preprocessing steps were applied. First, the old German data uses the “Alte Deutsche Rechtschreibung” whereas the newer parts of the corpus use the “Neue Deutsche Rechtschreibung”. We tried to normalize the text by converting the whole text to the “Neue Deutsche Rechtschreibung”. In a first step, we search for words that are only correct according to the old writing rules. Therefore, we selected all words in the corpus, that are correct according to the hunspell lexicon using the old rules, but not correct according to the hunspell lexicon using the new rules. In a second step we tried to find the correct spelling according to the new rules. Therefore, we first applied some rules describing how words changed from one spelling system to the other like replacing ’ß’ by ’ss’. If the new word is a correct word according to the hunspell lexicon using the new spelling rules, we map the words.

If translating from German to French or English, we apply compound splitting as described in (Koehn and Knight, 2003) to the German corpus.

As a last preprocessing step we remove sentences that are too long and empty lines to get the final corpus.

3 Word Reordering Model

Reordering was applied on the source side prior to decoding through the generation of lattices encoding possible reorderings of each source sentence. These possible reorderings were learned based on the POS of the source language words in the training corpus and the information about alignments between source and target language words in the corpus (Niehues and Kolss, 2009),(Rottmann and Vogel, 2007). Depending on the language pair, different types of reordering rules are applied. When translating from English to German most of the changes in word order consist of a shift to the right while typical word shifts in German to English translations take place in the reverse direction.

4 Translation Model

We apply Phrase Table Smoothing as described in (Jan Niehues and Waibel, 2009) in all systems.

4.1 Word Alignment

In most phrase-based SMT system the heuristic grow-diag-final-and is used to combine the alignments generated by GIZA++ from both directions. Then these alignments are used to extract the phrase pairs.

We used a discriminative word alignment model (*DWA*) to generate the alignments as described in Niehues and Vogel (2008) instead and then used the same method to extract the phrase pairs. This model is trained on a small amount of hand-aligned data and uses the lexical probability as well as the fertilities generated by the GIZA++ Toolkit and POS information. We used all local features, the GIZA and indicator fertility features as well as first order features for 6 directions. The model was trained in three steps, first using the maximum likelihood optimization and afterwards it was optimized towards the alignment error rate. For more details see Niehues and Vogel (2008).

Guzman et al. reported in (2009) improvements by adding a feature indicating the number of unaligned words on the source side and the same feature for the target side. For the system translating from German to English we also integrated this feature.

4.2 Bilingual language model

Since, for example, Allauzen et al. reported in (2009) improvemnts using their N-code translation system, we tried to integrate a billingual language model into the phrase-based translation system.

In our approach a token in the billingual language model consits a target word and all source words, it is alinged to. The tokens are ordered according to the target word order. In the example **??**, this would lead to the text: *I_Ich went_bin_gegangen home_Hause*.

As shown in the example one problem with this approach is, that unaligned source words are ignored. This could be overcome by using a second language model in the inverse direction. But since then, the sentence would not be build from right to left, the integration into the decoder is more complex and this first approach we only used a language model based on the target word order.

4.3 Bilingual POS language model

If we want to integrate POS-based phrase pairs into the decoder, we would like to be able to use longer POS-based phrase pairs, then word based phrase pairs. Otherwise, we can not make usage of the main advantage of POS-based phrase pairs, that there the data-sparness problem is less and we therefore can consider a bigger context.

But in most phrase-based decoders can not transate using phrase-pairs of different length for POS and word based phrase pairs. If we instead use a billinigual POS-based language model, the context of both language models are idepented of each other. Consequently, autmatically, longer context for the POS-based langauge model will be selected.

4.4 Lattice Phrase Extraction

For the test sentences the POS-based reordering allows us to change the word order in the source sentence, so that the sentence can be translated more easily. But this approach does not reorder the training sentences. This may cause problems for phrase extraction, especially for long-range reorderings. For example, if the English verb is aligned to both parts of the German verb, this phrase can not be extracted, since it is not continuous on the German side. In the case of German as source language, the phrase could be extracted if we also reorder the training corpus.

Therefore, as described in Jan Niehues and Waibel (2009) we build lattices that encode the different reorderings for every training sentence. Then we can not only extract phrase pairs from the monotone source path, but also from the reordered paths. So it would be possible to extract the example mentioned before, if both parts of the verb were put together by a reordering rule. To limit the number of extracted phrase pairs, we extract a source phrase only once per sentence even if it may be found on different paths. Furthermore, we do not use the weights in the lattice.

If we use the same rules as for the test sets, the lattice would be so big that the number of extracted phrase pairs would be still too high. As mentioned before, the word reordering is mainly a problem at the phrase extraction stage if one word is aligned to two words which are far away from each other in the sentence. Therefore, the short-range reordering rules do not help much in this case. So, only the long-range reordering rules were used to generate the lattice for the training corpus.

5 Results

We submitted translations for the English-German and German-English for the Shared Translation Task. In the following we present the experiments we conducted for both translation directions applying the aforementioned models and extensions to the baseline systems. The performance of each individual system configuration was measured applying the BLEU metric. All BLEU scores are calculated on the lower-cased translation hypotheses. The individual systems that were used to create the submission are indicated in bold.

5.1 English-German

The baseline system for English-German applies short-range reordering rules and phrase table smoothing. The language model is trained on the News corpus. By expanding the coverage of the rules to enable long-range reordering the score on the test set could be slightly improved. We then combined the target language part of the Europarl and NewsCommentary corpora with the News corpus to build a bigger language model which resulted in an increase of 0.11 BLEU points on the development set and an increase of 0.25 points on the test set. Applying the bilingual Word language model as described above led to 0.04 points improvement on the test set and a very slight improvement on the development set.

Table 1: Translation results for English-German (BLEU Score)

System	Dev	Test
Baseline	15.30	15.40
+ Long-range Reordering 15.25	15.44	
+ EPNC LM	15.36	15.69
+ bilingual Word LM	15.37	**15.73**
+ bilingual POS LM	15.42	15.67
+ unaligned Word Feature	15.65	15.66
+ bilingual Stem LM	15.57	15.74

This system was used to create the submission to the Shared Translation Task of the WMT 2010. However, after submission we performed additional experiments which could further improve our results as follows. Adding the bilingual POS language model decreased the score on the test set while the result on the development set could be slightly improved. A similar observation can be made when introducing an additional feature to the phrase table indicating the amount of unaligned words within each phrase pair. Although this feature improved quite a lot on the development set (0.23 BLEU points), the score on the test set remains practically unchanged. However, when adding a third bilingual language model based on stem information, the highest score of 15.74 on the test set could be reached. Although the score on the development set slightly decreased with respect to the previous experiment, nonetheless an increase with respect to the submission system is observable.

5.2 German-English

For the German to English translation system, the baseline system does already use short range reordering rules and the discriminative word alignment. This system does use only the language model trained on the news corpus. By adding the possibility to model long-range reorderings based on POS-based rules, we could improve the system by 0.6 BLEU points. Adding the big language model using also the english gigaword coprus we could improve by 0.3 BLEU points. We got an additional improvement by 0.1 BLEU points by adding the lattice phrase extraction.

Both, the word-based and POS-based billingual language, could improve the translation quality measure in BLEU. Together they improved the system performance by 0.2 BLEU points.

The best results could be achive by using also the unalinged word feature for source and target leading to the best preformance on the test set of 22.09.

Table 2: Translation results for German-English (BLEU Score)

System	Dev	Test
Baseline	20.94	20.83
+ Long-range reordering	21.52	21.43
+ Giga Word LM	21.90	21.71
+ Lattice phrase extraction	21.94	21.81
+ billingual Word LM	21.94	21.87
+ billingual POS LM	22.02	22.05
+ unaligned Word Feature	22.09	22.09

6 Conclusions

References

Alexandre Allauzen, Josep Crego, Aurélien Max, and François Yvon. 2009. LIMSI's statistical translation system for WMT'09. In *Fourth Workshop on Statistical Machine Translation (WMT 2009)*, Athens, Greece.

Francisco Guzman, Qin Gao, and Stephan Vogel. 2009. Reassessment of the Role of Phrase Extraction in PBSMT. In *MT Summit XII*, Ottawa, Ontario, Canada.

Muntsin Kolss Jan Niehues, Teresa Herrmann and Alex Waibel. 2009. The Universität Karlsruhe Translation System for the EACL-WMT 2009. In *Fourth Workshop on Statistical Machine Translation (WMT 2009)*, Athens, Greece.

Philipp Koehn and Kevin Knight. 2003. Empirical Methods for Compound Splitting. In *EACL*, Budapest, Hungary.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *ACL 2007, Demonstration Session*, Prague, Czech Republic, June 23.

Jan Niehues and Muntsin Kolss. 2009. A POS-Based Model for Long-Range Reorderings in SMT. In *Fourth Workshop on Statistical Machine Translation (WMT 2009)*, Athens, Greece.

Jan Niehues and Stephan Vogel. 2008. Discriminative Word Alignment via Alignment Matrix Modeling. In *Proc. of Third ACL Workshop on Statistical Machine Translation*, Columbus, USA.

Kay Rottmann and Stephan Vogel. 2007. Word Reordering in Statistical Machine Translation with a POS-Based Distortion Model. In *TMI*, Skövde, Sweden.

Ashish Venugopal, Andreas Zollman, and Alex Waibel. 2005. Training and Evaluation Error Minimization Rules for Statistical Machine Translation. In *Workshop on Data-drive Machine Translation and Beyond (WPT-05)*, Ann Arbor, MI.

Stephan Vogel. 2003. SMT Decoder Dissected: Word Reordering. In *Int. Conf. on Natural Language Processing and Knowledge Engineering*, Beijing, China.