Statistical Structure in Language Processing
Phrase based models

Cristina Gârbacea

10407936

cr1st1na.garbacea@gmail.com

Sara Veldhoen

10545298

sara.veldhoen@student.uva.nl

Abstract

Bla bla bla

1 Introduction

In this paper we explore the utility of phrase based models inside a statistical machine translation system. As compared to our previous assignment where we only used word aligned models, in this assignment our goal is to build an efficient phrase pair extraction tool that would extract phrase pairs of up to length 4 from an aligned given parallel training corpus.

Since word based models translate words as atomic units, many statistical translation systems based on one-to-one correspondences between source and target words fall short of capturing dependencies between groups of words. This assumption is further restricted in cases where each source word is aligned to exactly one target word, as in IBM Model 1. Given this incapacity to capture context around a word and word order in general, phrase based models can overcome this limitation by treating phrases as atomic translation units and making use of local context in the translation process.

2 Phrase Extraction and weight estimation

In this section, we present our approach to the extraction of phrase pairs from the corpus. Subsequently, we

Phrase Extraction The number of possible phrase pairs per sentence pair is huge: each sentence can be partitioned in a vast amount of ways, and each partition could form a phrase pair with any partition in the paired sentence.

In order to reduce the space, we consider only phrase pairs that are consistent with the alignments produced by IBM models. As in (1), consistency

is defined as follows:

$\langle \bar{e}, \bar{f} \rangle$ is consistent with $A \Leftrightarrow$

$$\forall e_i \in \bar{e} : \langle e_i, f_j \rangle \in A \Rightarrow f_j \in \bar{f},$$
$$\text{and } \forall f_j \in \bar{f} : \langle e_i, f_j \rangle \in A \Rightarrow e_i \in \bar{e},$$
$$\text{and } \exists e_i \in \bar{e}, f_j \in \bar{f} : \langle e_i, f_j \rangle \in A.$$

For this assignment, the symmetrized alignments of the corpus sentences were given. We base our extraction algorithm on the one presented in (1, page 133). We iterate over all windows up to a certain length in the English sentence, and find the foreign windows that are consistent given the alignment. For all valid pairs of windows, we extract the corresponding phrase pair.

Conditional Probability Estimates After having extracted the phrase pairs, we compute the conditional translation probability estimates for a foreign phrase $\overline{f}$ given an English phrase $\bar{e}$, using the following formula:

$$\phi(\overline{f}|\bar{e}) = \frac{count(\bar{e}, \overline{f})}{\sum_{\overline{f}_i} count(\bar{e}, \overline{f}_i)}$$

Here $count(\bar{e}, \overline{f})$ denotes in how many sentence pairs a specific phrase occurs and is extracted. To get relative frequency estimates, we normalize this value by the count of occurences of all phrase pairs containing the English phrase $\bar{e}$ inside the whole corpus.

Joint Probability Estimates In (2) quite a different approach is taken to phrase based translation. The idea of a noisy channel, that a foreign sentence is a corrupted version of an original English sentence, is abandoned. Rather, the two sentences are considered different substantiation of a bag of concepts. In this framework, the probability of a phrase pair is a joint probability conditioned

on a concept. In practice, we do not explicitly model the concept but view the phrase pair itself as a concept, so its weight is just the joint translation probability of the two phrases: $t(\bar{e}, \bar{f})$.

The estimation of the translation probabilities in (**?**) is done in an adapted version of expectation maximization. In the first step, high-frequency n-grams are determined in the bilingual corpus. The translation probabilities are initially estimated based on the length of the both phrases and the sentences in which they occur. Subsequently, fractional counts are collected based on the initial estimates.

In our implementation, we do not consider all alignments of all possible phrases, but instead base the extraction of phrases on the symmetrized word alignments from IBM models. Therefore, we can st

3   Moses

4   Bleu

5   Experiments and Results

6   Conclusion

References

[1] Philipp Koehn, 2010. *Statistical Machine Translation*. Cambridge University Press.

[2] Daniel Marcu and William Wong, 2002. *A phrase-based, joint probability model for statistical machine translation*. Association for Computational Linguistics.

[3] Franz Josef Och, Christoph Tillmann, and Hermann Ney, 1999. *Improved alignment models for statistical machine translation* Proceedinfs of the Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora.