Abstract

We present a context-sensitive chart pruning method for cxy-style MT decoding. Source phrases that are unlikely to have aligned target constituents are identified using sequence labelers learned from the parallel corpus, and speed-up is obtained by pruning corresponding chart cells. The proposed method is easy to implement, orthogonal to cubic pruning and additive to its pruning power. On a full-scale English-to-German experiment with a string-to-tree model, we obtain a speed-up of more than 60% over a strong baseline, with no loss in BLEU.

Pruning by Labelling

In contrast to monolingual parsing, our pruning decisions are based on the source context, its target translation and the mapping between the two. The key question is: how to inject target syntax and word alignment information into our labelling model? This is important since the syntactic correspondence between different language pairs is different. For example, “the products” does not have a consistent alignment on the target side (en-de), while it does on the right (en-jp).

Gold-standard Labelling

For each training sentence pair, gold-standard B-tags and e-tags are assigned separately to the source words. First, we initialize both tags of each source word to 0s. Then, we iterate through all target constituent spans, and for each span, we find its corresponding source phrase, as determined by the word alignment. If a constituent exists for the phrase pair, the b-tag of the first word and the e-tag of the last word in the source phrase are set to 1s, respectively.

Input forward alignment $A_{i,j}$, backward alignment $A_{i,j}^{\prime}$ and 1-best parse tree $T$ for $f$

Output Tag sequences $b$ and $e$ for $e$

1: procedure $T$($e$, $f$, $r$, $A$, $A'$)
2: $l = |e|$
3: for $i = 0$ to $l - 1$ do
4: \hspace{1em} $b[j] \leftarrow 0, e[j] \leftarrow 0$
5: for $j' = j$ to $i + 1$ do
6: \hspace{1em} $s \leftarrow \{A[k] \mid k \in \{j', j'\}]$
7: if $|s| \leq 1$ then continue
8: \hspace{1em} $i \leftarrow \min(i, j), j \leftarrow \max(s)$
9: if CONSISTENT($i, j, i', j'$) then
10: \hspace{1em} $b[j'] \leftarrow 1, e[j'] \leftarrow 1$
11: procedure CONSISTENT($i, j, i', j'$)
12: \hspace{1em} $t = \{A[k] \mid k \in \{i, i', j, j'\}\}$
13: return $\min(t) \geq \theta$ and $\max(t) \leq \theta'$

We use binary tags to indicate whether a source word can start or end a multi-word phrase that has a consistently aligned target constituent. Under this scheme, a b-tag value of 1 indicates that a source word can start or end a phrase of a consistently aligned target phrase; similarly an e-tag of 0 indicates that a word cannot end a source phrase. If either the b-tag or the e-tag of an input phrase is 0, the corresponding chart cells will be pruned. The pruning effects of the two types of tags are illustrated below. In general, 0-valued B-tags prune a whole column of chart cells and 0-valued e-tags prune a whole diagonal of cells; and the chart cells on the first row and the top-most cell are always kept so that complete translations can always be found.

We build a separate labeler for each tag type using gold-standard b- and e-tags, respectively. The labelers are trained with maximum-entropy models (Curran and Clark, 2003; Ratinaparkhi, 1996), using features similar to those used for supertagging for CCG parsing (Clark and Curran, 2004). During testing, in order to prevent overpruning, a tag value of 0 is assigned to a word only if its marginal probability is greater than a cut-off value $\theta$.

Method Overview

We study a chart pruning method for cxy-style MT decoding that is orthogonal to cubic pruning and additive to its pruning power. The main intuition of our method is to find those source phrases (i.e. any sequence of consecutive words) that are unlikely to have any consistently aligned target counterparts according to the source context and grammar constraints.

We call our method context-sensitive pruning (CSP); it can be viewed as a bilingual adaptation of similar methods in monolingual parsing (Roark and Hollingshead, 2008; Zhang et al., 2010) which improve parsing efficiency by “closing” chart cells using binary classifiers. Our contribution is that we demonstrate such methods can be applied to synchronous-grammar parsing by labelling the source-side alone. This is achieved through a novel training scheme where the labelling models are trained over the word-aligned best and gold-standard pruning labels are obtained by projecting target-side constituents to the source words. To our knowledge, this is the first work to apply this technique to MT decoding.

Results

A WMT Moses string-to-tree system is used as our baseline and decoding speed is measured by the average decoding time and average number of hypotheses generated per sentence. As shown below, the CSP decoder, which considers far fewer chart cells and generates significantly fewer sub-translations, consistently outperforms the slower baseline. It ultimately achieves a BLEU score of 1.86 at a probability cutoff value of 0.98, slightly higher than the highest score of the baseline.

References