Lexicalized Beam Thresholding Parsing with Prior and Boundary Estimates *

Deyi Xiong[1,2], Qun Liu[1], and Shouxun Lin[1]

[1] Institute of Computing Technology, Chinese Academy of Sciences
P.O. Box 2704, Beijing 100080, China
{dyxiong, liuqun, sxlin}@ict.ac.cn
[2] Graduate School of Chinese Academy of Sciences

Abstract. We use prior and boundary estimates as the approximation of outside probability and establish our beam thresholding strategies based on these estimates. Lexical items, e.g. head word and head tag, are also incorporated to lexicalized prior and boundary estimates. Experiments on the Penn Chinese Treebank show that beam thresholding with lexicalized prior works much better than that with unlexicalized prior. Differentiating completed edges from incomplete edges paves the way for using boundary estimates in the edge-based beam chart parsing. The beam thresholding based on lexicalized prior, combined with unlexicalized boundary, runs faster than that only with lexicalized prior by a factor of 1.5, at the same performance level.

1 Introduction

In the recent development of parsing technology, lexicalized grammars has been used in several state-of-the-art parsers(see [1][2] etc.) to pursue high accuracy because they control not only structural dependencies, but also lexical dependencies, lexico-structural dependencies(see [3]). In this paper, we just consider lexicalized context-free grammars(LCFG). LCFG is a CFG with its nonterminals lexicalized by some lexical items(see [4]). For example, in Collins' bilexical grammars, each nonterminal is associated with a word(called the head of the corresponding constituent) and a POS tag of the head.

When CKY chart parsing techniques are used to bilexical context-free grammars, the time complexity is not $O(n^3)$, but $O(n^5)$. A CKY chart parser can be considered as a two-dimensional matrix of cells. In each chart cell, there are $O(n)$ edges because of the $O(n)$ possible choices for head words to be associated with nonterminals of edges. When fundamental rule is used between two neighbor cells, the algorithm requires additional time $O(n^2)$.

Because of the heavy work load for lexicalized parsers, edge pruning techniques, like beam thresholding, are usually used by practical parsing algorithms. The key problem for beam thresholding is how to select a evaluation function

* This work was supported in part by National High Technology Research and Development Program under grant #2001AA114010