

A Probabilistic Chart Parser Implemented with an Evolutionary Algorithm ^{*}

L. Araujo

Dpto. Sistemas Informáticos y Programación. Universidad Complutense de Madrid.
Spain. lurdes@sip.ucm.es

Abstract. Classic parsing methods use complete search techniques to find the different interpretations of a sentence. However, the size of the search space increases exponentially with the length of the sentence or text to be parsed and the size of the grammar, so that exhaustive search methods can fail to reach a solution in a reasonable time. Nevertheless, large problems can be solved approximately by some kind of stochastic techniques, which do not guarantee the optimum value, but allow adjusting the probability of error by increasing the number of points explored. Evolutionary Algorithms are among such techniques. This paper presents a stochastic chart parser based on an evolutionary algorithm which works with a population of partial parsings. The paper describes the relationships between the elements of a classic chart parser and those of the evolutionary algorithm. The model has been implemented, and the results obtained for texts extracted from the Susanne corpus are presented.

keywords: Evolutionary programming, Partial Parsing, Probabilistic Grammars

1 Introduction

Parsing a sentence can be sought as a procedure that searches for different ways of combining grammatical rules to find a combination which could be the structure of the sentence. A bottom-up parser starts with the sequence of lexical classes of the words and its basic operation is to take a sequence of symbols to match it to the right-hand side of the rules. Thus, this parser can be implemented simply as a search procedure for this matching process. However, such implementation would be extremely expensive because the parser would try the same matches again and again. This problem is avoided in the chart parsing algorithms by introducing a data structure called *chart* [1]. This structure stores the partial results of the matchings already done.

Classical parsing methods are based on complete search techniques to find the different interpretations of a sentence. However, experiments on human parsing suggest that people do not perform a complete search of the grammar while

^{*} Supported by projects TIC2003-09481-C04 and 07T/0030/2003.