Analogical Modeling with Bias — allowing Feedback and Centering

Christer Johansson and Lars G. Johnsen

Dept. of Linguistics and Literature
University of Bergen
N-5007 Bergen, Norway
{christer.johansson, lars.johnsen}@lili.uib.no

Abstract. We show a computationally efficient approximation (cf. [1]) of a full analogy model [2, 3], implemented in a computer program, and tested on the CoNLL2000 chunk tagging task [4], putting clause boundaries around mainly np and vp phrases. Our implementation showed to be competitive with other memory based learners. It deviates only slightly from the theoretical model. First, it implements a version of homogeneity check, which does not account fully for nondeterministic homogeneity. Second, it allows feedback of the last classification, and thirdly it allows centering on some central feature positions. Positions containing a) those parts-of-speech tags and b) those words that are to be given a chunk tag are given a weight which is given by how many match patterns that are equally or more general. A match on two centered features gives its patterns an extra weight given by the number of features. The results can be summarized as follows: a) using only lexical features performs below baseline. b) The implementation without anything extra, performs as the baseline for five parts-of-speech features, and centering improves the results. c) Feedback on its own does not improve results, while feedback + centering improves results more than just centering. Feedback on its own makes results deteriorate. The results exceed F=92, which is comparable with some of the best reported results for Memory Based Learning on the chunk tagging task.

1 Introduction

Analogical modeling (AM) is a (memory based) method to evaluate the analogical support for a classification [2, 3, 5]. Chandler [6] suggested AM as an alternative to both rule based and connectionist models of language processing and acquisition. AM defines a natural statistic, which can be implemented by comparisons of subsets of linguistic variables, without numerical calculations [5]. The natural statistic works as a selection mechanism, selecting those patterns in the database which most clearly points out a class for a novel pattern.

The original AM model compares all subsets of investigated variables. This may cause an exponential explosion in the number of comparisons, which has made it difficult to investigate large models with many variables (> 10) combined