

# Assigning Function Tags with a Simple Model

Vasile Rus and Kirtan Desai

Department of Computer Science  
The University of Memphis  
Institute for Intelligent Systems  
Fedex Institute of Technology  
Memphis, TN 38120, USA  
vrus@memphis.edu

**Abstract.** This paper presents a method to assign function tags based on a Naive Bayes approach. The method takes as input a parse tree and labels certain constituents with a set of functional marks such as logical subject, predicate, etc. The performance reported is promising, given the simplicity of a Naive Bayes approach, when compared with similar work.

## 1 Introduction

Syntactic structure is an important phase in a variety of language tasks since it provides important information for semantic interpretation. State-of-the-art statistical parsers, the tools that generate syntactic structures, are freely available nowadays but their output is limited to basic structures and are not able to deliver richer syntactic information such as logical subject or predicate. Most of the available statistical parsers are trained on Penn Treebank [7] and are only able to identify simple phrases such as NP, VP or S although Penn Treebank contains function tags and remote dependencies coded as traces. This paper presents a naive Bayes approach to augment the output of Treebank-style syntactic parsers with functional information.

In Section 2.2 of Bracketing Guidelines for Treebank II [7], there are 20 function tags grouped in four categories: form/function discrepancies, grammatical role, adverbials, and miscellaneous. Up to 4 function tags can be added to the standard syntactic label (NP, ADVP, PP, etc.) of each bracket. Those tags were necessary to distinguish words or phrases that belong to one syntactic category and is used for some other function or when it plays a role that is not easily identified without special annotation. We rearrange the four categories into four new categories based on corpus evidence, in a way similar to [1]. The new four categories are given in Table 1 and were derived so that no two labels from same new category can be attached to the same bracket.

We present in this paper a naive Bayes approach to build a system that automatically assigns function tags to constituents in parse trees. The function tags assignment problem is viewed as a classification problem, where the task is to select the correct tag from a list of candidate tags. The results are reported per category based on the new categories mentioned above.