Final Paper: Language models BSc AI 2013

Eszter Fodor, Sharon Gieske
(5873320, 6167667)

May 20, 2013

Abstract

This paper discribes our implementation of a Markov POS tagger, trained and tested on various sections of the Wall Street Journal corpus. We will describe the research question that we wanted to answer, how we approached the problem and what our results were.

Introduction

The problem that needed to be solved during the four assignments was how to calculate the most probable Part-Of-Speach (POS) tag sequence using tagging and smoothing on a training corpus. The program that needed to be written would eventually be able to tag the test corpus with POS tags.

Some POS-tag sequences are less possible than others which information can be used to analyze texts to learn whether it is grammatically correct or not. Another importance of POS-tagging is to disambiguate words within a sentence. This can be done by taking the previous words and their POS-tags into account. A program that is able to tag sentences accurately can be used on various corpora to learn what kind of sentences it consists of.

Approach

Results

Conclusion/Discussion