عنوان مقاله:

Feature Engineering in Persian Dependency Parser

نویسندگان:

.S. Lazemi - *Department of Computer Eng., University of Kashan, Kashan, Iran*

.H. Ebrahimpour-komleh - *Department of Computer Eng., University of Kashan, Kashan, Iran*

خلاصه مقاله:

Dependency parser is one of the most important fundamental tools in the natural language processing, which extracts structure of sentences and determines the relations between words based on the dependency grammar. The dependency parser is proper for free order languages, such as Persian. In this paper, data-driven dependency parser has been developed with the help of phrase-structure parser for Persian. The defined feature space in each parser is one of the important factors in its success. Our goal is to generate and extract appropriate features to dependency parsing of Persian sentences. To achieve this goal, new semantic and syntactic features have been defined and added to the MSTParser by stacking method. Semantic features are obtained by using word clustering algorithms based on syntagmatic analysis and syntactic features are obtained by using the Persian phrase-structure parser and have been used as bit-string. Experiments have been done on the Persian Dependency Treebank (PerDT) and the Uppsala Persian Dependency Treebank (UPDT). The results indicate that the definition of new features improves the performance of the dependency parser for the Persian. The achieved unlabeled attachment score for PerDT and UPDT are 89.17% and 88.96% respectively.

کلمات کلیدی:

Dependency Parser, Phrase-structure parser, MSTParser, Stacking, Persian

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/894053