

Automatic Extraction of Case Frames for Chinese Verbs

XIAOHONG WU
WENLIANG CHEN
HITOSHI ISAHARA

National Institute of Information & Comm. Tech., Kyoto, Japan

ABSTRACT

Information on case frames of verbs is very useful for semantic and syntactic disambiguation needed in NLP, especially for parsing and MT where incorrect analysis or mismatching are frequently seen. However, extracting case frames for Chinese verbs is a difficult task as the Chinese language lacks obviously marked case frames. Prepositions could be alternatively considered as case markers; however, it is not as straightforward as we expect. Despite of this fact, there are still some structures in which the prepositions do play an important and special role in the syntactic constructions and can be considered as typical case markers, for example, the “ba” in ba-construction. In this paper we discuss the method that we employ to automatically extract case frames by using the prepositions as the key words as the first step for building a case frame dictionary. To get a good coverage as well as better and more reliable results we make use of large unlabelled corpora. We employ a dependency parser to parse the corpus, taking advantage of the dependency relationships provided by the parser to extract case frames. We lay focus on the kind of syntactic constructions in which the object is fronted by employing the prepositions and thus form a relatively fixed case frame. Or the prepositions are right after the verbs and also function as the object of the verbs. By defining strict rules using the context free grammar, we successfully extract the case frames of the verbs that can be used in such constructions and thereby get a list of verbs together with a list of nouns (its arguments) that can fill in such frames. This paper thus addresses the problems and solutions of such a task from a linguistic point of view, focusing on the methods which can facilitate the automatic extraction by the computer.