Building a Situation-based Language Knowledge Base

Qiang ZHOU, Zushun CHEN

State Key Laboratory of Intelligent Technology and Systems
Dept. of Computer Science and Technology
Tsinghua University, Beijing 100084, P. R. China
`zq-lxd@tsinghua.edu.cn`

Abstract. Language resources are very important for natural language processing research and applications. This paper will introduce our ongoing research work to build a situation-based language knowledge base for the Chinese language, based on two basic language resources: three Chinese semantic lexicons and a large scale Chinese treebank. We developed a supporting platform to make full use of the abundant information contained in current Chinese semantic lexicons so as to gradually summarize the complete situation descriptions, organize them as situation network and build corresponding descriptive definition dictionary for different concepts. We explored an efficient algorithm to link from syntax to semantics so as to introduce suitable semantic explanations into current Chinese treebank and gradually build a situation-based semantically-annotated corpus. All these research work will lay a good foundation for the computational infrastructure in Chinese natural language processing.

1 Introduction

Language resources are very important for natural language processing research and applications. In recent years many researchers have devoted themselves to the construction of large-scale language resources. Nowadays, there are two types of commonly used language resources. One is syntactically annotated corpora. Some typical examples include the Penn Treebank for English [9], the Prague Dependency Treebank for Czech [6] and the TIGER treebank for German [3]. The other is the semantic lexicons. Most of them are manually compiled by linguists or lexicographers. Some typical examples are the WordNet [10] and Levin's English verb classes [8]. The key issue is how to integrate these two types of language resource so as to build the linking bridge between syntax and semantics. The Proposition Bank (PropBank) [7] and FrameNet [1] projects have made some tentative explorations in these respects.

Unlike these research projects, we propose a new situation-based language knowledge description framework. In this framework, we use situation as a mathematical model to describe a cognition scheme and try to define a concept under its generating situation. Therefore, the situation theory can serve as a unified theoretical framework for constructing lexical semantics and the natural language knowledge infrastructure built upon it. This paper introduces our ongoing research work to build a situation-based language knowledge base for the Chinese language based on two basic language resources: three Chinese semantic lexicons and a large scale Chinese treebank.