

金弓冶（河原大輔准教授）

「High-quality Knowledge Acquisition of Predicate-argument Structures for Syntactic and Semantic Analysis」

（構文・意味解析のための高品質な述語項構造知識の獲得）

平成 28 年 3 月 23 日授与

In computer science, people have been trying hard to march toward one of the very ultimate goals which expects computers can truly understand human’s language. This task is named as text understanding. During the past decades, great efforts have been made to develop various kinds of techniques to achieve text understanding, which can further provide us with the space to discover more humanized ways for human-computer interaction. To complete this ultimate goal, it may involve multiple domains such as cognitive science, linguistics and even psychology. In Natural Language Processing (NLP), which is a field of computer science and artificial intelligence, for computers to understand text in human language, the priority is to discover and clarify the internal relations of the words within a sentence. In this thesis, it is considered to be crucial to acquire large-scale knowledge for improvement of varieties of NLP tasks. Thanks to data explosion, unprocessed data that potentially contain a large amount of useful information can be relatively easier to acquire, and thus it is promising to acquire knowledge from these raw texts using automatic approaches. Automatic knowledge acquisition is always dependent on the fundamental analysis such as syntactic parsing and semantic analysis. Without considering the automatic analyzing errors, the knowledge will be extremely noisy and may cause bad effects when we apply this kind of knowledge to NLP tasks. In this thesis, we solve the abovementioned problems by proposing a framework of acquiring high-quality knowledge from unlabeled raw text that contains less noise, and apply the knowledge to syntactic and semantic analysis.

In this thesis, we present a framework for knowledge construction from high-quality dependencies that are selected from automatic dependency parses. Figure 1 shows the overview of the proposed framework. For the raw text, we apply dependency parsing and acquire automatic dependency parses, which inevitably contain parsing errors. A high-quality dependency selection approach is proposed to filter these erroneous automatic parses and only select those with high reliability. From the high-quality dependency parses, we extract predicate-argument structures (PASs), which contain the abstract of a sentence. PASs can be used as additional knowledge for other NLP tasks. Afterwards, semantic clustering is applied to merge all the predicate-argument structures with similar meaning. The clustered PASs are called case frames, which are regarded as another type of knowledge. These are a type of semantic frames that distinguish between the different usages of each predicate. The PAS knowledge is then used to improve dependency parsing. We also use case frames to improve Semantic Role Labeling (SRL). We apply a similar selection approach to select high-quality semantic roles for deep case PAS extraction and deep case frame construction. Deep case frames are used for further SRL improvement. The experimental results show that the quality management of acquired knowledge is beneficial for various NLP tasks.

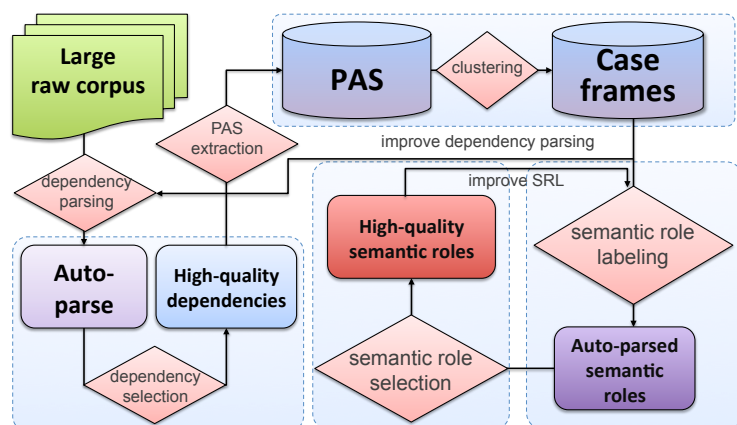


Figure 1: High-quality knowledge acquisition overview