CRII:III: Composing Process Knowledge using Semantic Roles

Niranjan Balasubramanian, Stony Brook University

Large-scale text resources and fact databases have spurred significant advances in AI systems that answer factual questions (e.g., "When was Bill Clinton born?"). To address more challenging problems that go beyond factual answer retrieval, AI systems need to be able to reason about general scenarios. They need to know general truths or generalities and how to apply them to specific situations. Acquiring this type of general knowledge and using it to reason effectively remains a huge challenge.

We explore this challenge for reasoning about *processes* in order to answer questions about specific scenarios involving them. In particular we propose to compose knowledge about processes in terms of semantic roles (e.g., What is undergoing the process? What is the result? etc.). We develop solutions to automatically construct a large repository of simple process knowledge, and demonstrate its use to answer process questions that go beyond fact lookup.

Existing lexical semantic resources provide similar forms of semantic knowledge about general open-domain actions but lack coverage on scientific processes. FrameNet, for instance, does not have entries for nearly half of the processes described in 4th grade science exams. The coverage is likely worse for higher grade levels with deeper knowledge domains.

In response we propose to investigate methods to automatically construct a large repository of simple semantic role based knowledge about processes. Adapting existing semantic role labeling systems to work well on new domains is difficult. We take a different approach. Our key premise is that rather than building a semantic role labeler that works well on any sentence, we proactively gather sentences that convey information in expected constructs. We propose a framework that combines techniques from extraction and joint inference for finding semantic roles and iteratively expands its knowledge to discover roles on its own. Specifically, we will make three main contributions:

- The first comprehensive, large-scale **knowledge base of processes** in the grade science domain, describing the roles and changes involved in that process.
- Methods for **automatic extraction of process knowledge** using information extraction and joint inference.
- A framework for **iterative knowledge expansion**, which allows the system to discover new roles involved in a process and expand the process representation to accommodate them.

Intellectual Merit

This work investigates a new direction in acquiring semantic roles by targeting sentences that express information in expected ways. The work will lead to better understanding and advancement of cross-sentence alignment of semantic roles and collective labeling. This work will push understanding on continuous learning (similar to the NELL project for relation extraction) with the iterative expansion methods and increase understanding of the connection between regularities of syntactic realizations and semantic roles. Overall the project also contributes to advances in leveraging unlabeled data and methods for balancing representational needs against extraction capabilities.

Broader Impact

Automatic knowledge extraction is fundamental to advancing Artificial Intelligence. This work contributes towards building systems capable of understanding knowledge in texts and reason with them. Better reasoning systems help information access and reduce information overload thereby accelerating research and discovery, as well as serve the information needs of the population at large.

Keywords Knowledge extraction, Semantic Role Labeling, Information Extraction, Question Answering