

Scalable Construction and Reasoning of Massive Knowledge Bases

Xiang Ren, Nanyun Peng and William Yang Wang

(Submission #48)

Summary

In today's information-based society, there is abundant knowledge out there carried in the form of natural language texts (e.g., news articles, social media posts, scientific publications), which spans across various domains (e.g., corporate documents, advertisements, legal acts, medical reports), which grows at an astonishing rate. Yet this knowledge is mostly inaccessible to computers and overwhelming for human experts to absorb. How to turn such massive and unstructured text data into structured, actionable knowledge, and furthermore, how to teach machines learn to reason and complete the extracted knowledge is a grand challenge to the research community.

Traditional IE systems assume abundant human annotations for training high quality machine learning models, which is impractical when trying to deploy IE systems to a broad range of domains, settings and languages.

In the first part of the tutorial, we introduce how to extract structured facts (i.e., entities and their relations for types of interest) from text corpora to construct knowledge bases, with a focus on methods that are weakly-supervised and domain-independent for timely knowledge base construction across various application domains.

In the second part, we introduce how to leverage other knowledge, such as the distributional statistics of characters and words, the annotations for other tasks and other domains, and the linguistics and problem structures, to combat the problem of inadequate supervision, and conduct low-resource information extraction.

In the third part, we describe recent advances in knowledge base reasoning. We start with the gentle introduction to the literature, focusing on path-based and embedding based methods. We then describe DeepPath, a recent attempt of using deep reinforcement learning to combine the best of both worlds for knowledge base reasoning.

Categories

Preferred Conference: *NAACL 2018*
2nd Choice Preference: *EMNLP 2018*

Download

[\[Paper \(PDF\)\]](#)