

Constructing Knowledge Nets from Real-world Natural Language Texts

Wei Liu
School of Computer Science and Software Engineering
The University of Western Australia
35 Stirling Highway, Crawley WA 6009, Australia
wei.liu@uwa.edu.au

Abstract

Mining knowledge from textual data has traditionally been applied on web-based publicly available resources, such as Wikipedia, online news, scientific publications and social medial such as Facebook and Twitter. In recent years, the increased maturity and accessibility of natural language processing and text mining algorithms have now raised serious interests beyond academia to uncover the chunk of corporate knowledge buried in hard-to-process natural language descriptions and reports. Such textual data are often in larger volumes and contain more valuable than numerical data.

To make unstructured textual data written in natural languages ready for use in downstream data mining tasks, a typical workflow consists of anonymisation, text normalisation, domain entity and relation recognition. Despite intensive research efforts in the past decade, this seemingly simple workflow remains challenging when facing real-world text. The dominant issues are but not limited to:

- Lack of labelled training data: despite the fact that supervised learning based algorithms are more effective than unsupervised ones, it is often difficult if not impossible to obtain large amount of labelled documents.
- Not-so-reproducible results: the results of current algorithms are highly dependent on the chosen combination of the pre-processing steps and their ordering. The performance are sensitive to the statistical and linguistic features chosen.
- Lack of domain-independent representations: there is the lack of assurance that algorithms performing well for one text will also work for another.

In this invited talk, we will take constructing knowledge nets as the objective, looking into the methods and techniques ranging from traditional feature-based distributional representation of words and phrases to the more recent deep learning enabled distributed vector representations. A demo

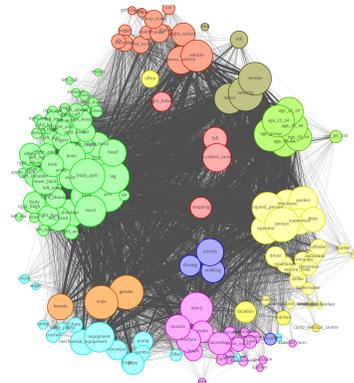


Figure 1: A Knowledge Net Constructed from Textual Records

of a prototype system (as shown in Figure 1) will showcase how the knowledge net construction workflow could be realised using textual records in real world databases.

Biography

Dr Wei Liu is a high-impact leading researcher in ontology learning from textual data. To date, her seminal work in 2012, “Ontology learning from text: A look back and into the future”, has received 190 citations. Her recent publications on using deep learning techniques for representing words and phrases are also well received in the community with best paper awards.

The research techniques and tools developed by her team are mostly domain agnostic, which builds a solid theoretical foundation for the knowledge mining and knowledge net traversal. Dr Liu also has extensive experience working on projects focusing on domain specific knowledge discovery and visualisation, with partners including mining companies, Public Transport Authority, Government agencies for resources and mining, and Medical Practitioners. These allow for accessing documents from a wide variety of domains and motivating algorithm developments to meet diverse challenges of real-world applications. Most of her recent work are developed on the Australian National NECTAR Research Cloud, testing out scalability with large datasets in order to prepare the software for deployment on commercial cloud service providers. She has up-to-date knowledge and experiences in building cloud-based solutions using AWS, Microsoft Azure and IBM Bluemix.

©2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC BY 4.0 License.
WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
<http://dx.doi.org/10.1145/3041021.3055360>

