A multi-agent framework for mining semantic relations from Linked Data

Hua-jun CHEN1†, Tong YU1†, Qing-zhao ZHENG1, Pei-qin GU1, Yu ZHANG2

(1) School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
(2) College of Information, Zhejiang Sci-Tech University, Hangzhou 310018, China
† E-mail: {huajunsir, ytcs}@zju.edu.cn

Received Aug. 13, 2011; Revision accepted Feb. 13, 2012; Crosschecked Feb. 27, 2012

Abstract: Linked data is a decentralized space of interlinked Resource Description Framework (RDF) graphs that are published, accessed, and manipulated by a multitude of Web agents. Here, we present a multi-agent framework for mining hypothetical semantic relations from linked data, in which the discovery, management, and validation of relations can be carried out independently by different agents. These agents collaborate in relation mining by publishing and exchanging inter-dependent knowledge elements, e.g., hypotheses, evidence, and proofs, giving rise to an evidentiary network that connects and ranks diverse knowledge elements. Simulation results show that the framework is scalable in a multi-agent environment. Real-world applications show that the framework is suitable for interdisciplinary and collaborative relation discovery tasks in social domains.

Key words: Semantic Web, Linked open data, Semantic association discovery

1 Introduction

Berners-Lee et al. (2001) envisioned the Semantic Web (SW) as a Web of data that is meaningful and understandable to any computer. The Web of data, when fully realized, will enable us to share structured data (e.g., spreadsheets and databases) as easily as we share documents, photos, and videos today. Conceptually speaking, the Web of data can be viewed as a graph layer that emerges on top of the current Web (Ayers, 2008). According to Berners-Lee et al. (2008), the Web of data has two faces: (1) the ‘Graph of Things’ (also called the Giant Global Graph, GGG), which encapsulates the semantic relations under investigation, with nodes representing concepts and edges representing relations that are annotated with evidence, and (2) the ‘Web of documents’, which contains a set of interlinked documents that serve as evidence. The Semantic Web will ignite a revolution of intelligent agents, which operate directly on the ‘Graph of Things’ and collaborate with each other to solve complex problems and accomplish intelligent tasks (Berners-Lee et al., 2001; Hendler, 2001; 2007). This technical trend will lead to the emergence of intelligent applications that take advantage of the Web of data to augment the underlying Web system’s functionalities, such as information retrieval and knowledge sharing (Mukherjea, 2005).

The concept of the Semantic Web is best manifested in the prosperous movement of Linked Data, which was initiated by Berners-Lee (2006). According to Heath and Bizer (2011), Linked Data “provides a publishing paradigm in which not only documents, but also data, can be a first class citizen of the Web, thereby enabling the extension of the Web with
a global data space based on open standards—the Web of data”. Inspired by Linked Data principles, the Semantic Web community has published a large set of datasets, covering a broad range of domains such as life sciences, healthcare, social networking, and e-commerce (Bizer et al., 2009). With the rapid growth of datasets on the Web, how to distill knowledge and insights from this wealth of data becomes an important problem.

In this paper, we focus on a particular knowledge discovery problem called relation discovery (also called link discovery, link prediction, relation mining, etc.), which typically means to find interesting relations (expressed as meaningful paths, subgraphs, patterns, etc.) from large-scale datasets (Tarjan, 1981; Deerwester et al., 1990; de Raedt et al., 2007). The Web of data, which is essentially a graphical data model, has provided excellent vehicle for the representation, mapping, and analysis of complex relations. With this background, semantic association discovery (SAD) is proposed to infer implicit or latent relations between arbitrary resources based on patterns discovered from the Web of data (Anyanwu and Sheth, 2003; Aleman-Meza, 2005; Anyanwu, 2007; Anyanwu et al., 2007).

Here, we illustrate the usefulness of SAD through a motivating story of mining social networks (Mika, 2005; Aleman-Meza et al., 2006). As shown in Fig. 1, the news contents from Web pages are important sources for mining social relations between public figures. The extracted relations, however, typically lack accurate semantic labels. Linked data can be used to discover direct or indirect evidence that annotates the extracted relations. For example, the frequent co-occurrences of ‘Obama’ and ‘Michelle Robinson’ can be annotated with the triple ⟨Obama, spouse, Michelle Robinson⟩ that can be queried from linked data.

This story illustrates a genre of ‘connect-the-dots’ applications, in which knowledge analysts typically use analytical tools to gather a set of interlinked intelligence resources to discern hidden and important relations, which often involves cross-domain knowledge integration and collaboration. A ‘connect-the-dots’ application requires that the ‘Graph of Things’ should be extracted from the ‘Web of documents’, which is then navigable and editable by multiple parties, and also capable of answering domain-specific complex problems.

To support the above genre of applications, semantic relations should be derived from intelligence resources and aggregated into a graph while connecting to their evidence for justification and validation. Accordingly, we propose a hypothesis-driven...