

Chapter 4

A Hybrid Concept Learning Approach to Ontology Enrichment

Zenun Kastrati

Norwegian University of Science and Technology (NTNU), Norway

Ali Shariq Imran

Norwegian University of Science and Technology (NTNU), Norway

Sule Yildirim Yayilgan

Norwegian University of Science and Technology (NTNU), Norway

ABSTRACT

The wide use of ontology in different applications has resulted in a plethora of automatic approaches for population and enrichment of an ontology. Ontology enrichment is an iterative process where the existing ontology is continuously updated with new concepts. A key aspect in ontology enrichment process is the concept learning approach. A learning approach can be a linguistic-based, statistical-based, or hybrid-based that employs both linguistic as well as statistical-based learning approaches. This chapter presents a concept enrichment model that combines contextual and semantic information of terms. The proposed model called SEMCON employs a hybrid concept learning approach utilizing functionalities from statistical and linguistic ontology learning techniques. The model introduced for the first time two statistical features that have shown to improve the overall score ranking of highly relevant terms for concept enrichment. The chapter also gives some recommendations and possible future research directions based on the discussion in following sections.

INTRODUCTION

Domain ontologies are a good starting point to model in a formal way the basic vocabulary of a given domain. They provide a broad coverage of concepts and their relationships within a domain. However, in-depth coverage of concepts is often not available, thereby limiting their use in specialized subdomain applications. It is also the business dynamics and changes in the operating environment which require

DOI: 10.4018/978-1-5225-5042-6.ch004

modification to an ontology (McGuinness, 2000). Therefore, the techniques for modifying ontologies, i.e. ontology enrichment, have emerged as an essential prerequisite for ontology-based applications.

An ontology can be enriched with lexical data either by populating the ontology with lexical entries or by adding terms to ontology concepts. The former means updating the existing ontology with new concepts along with their ontological relations and types. This increases the size of the existing ontology which requires more computational resources and more time to compute. Thus, making it less cost effective. The latter means adding new concepts without taking into account the ontological relations and types between concepts. Because of this, the ontology structure will remain the same but its concepts will be enriched with their synonym terms or linguistic variants.

Enrichment of ontology concepts is aiming at improving an existing ontology with new concepts. It is part of the iterative ontology engineering process (Faatz & Steinmetz, 2005). The core of this process is the learning approach which constitute tasks such as identification and acquisition of the relevant terminology through exploring various knowledge resources, and the creation of the concepts.

There is a variety of concept learning approaches that are available to enrich concepts of an ontology. These approaches rely on either linguistic, statistical, or hybrid techniques (Drumond & Girardi, 2008; Hazman, El-Beltagy, & Rafea, 2011). Although, these approaches proved useful for enriching ontologies of many domains, they do have some limitations, especially when it comes to semantic information of terms. The existing approaches use only contextual information without considering the semantic information of terms. Moreover, the contextual information is simply derived by distributional property of terms such as term frequency tf or term frequency inverse document frequency $tf*idf$, and co-occurrences of terms.

The focus of this chapter is to enlighten the reader with the ontology concept enrichment process, explore state-of-the-art methods and techniques in this regard, review input data resources, learning approaches and systems build upon them, discuss their limitations and to propose solutions and to give some recommendations accordingly. It also describes the SEMCON model to enriching the domain ontology with new concepts by combining contextual as well as the semantics of terms.

SEMCON uses unstructured data as input for ontology learning process and is composed of two parts - contextual and semantic. Context is defined as the part of a text or statement – passage that surrounds a given term and it determines term meaning. In this work, it is the cosine distance between the feature vectors of any two terms. The feature vectors are composed of values computed by both the frequency of occurrence of terms in corresponding passages, and the statistical features such as font type and font size. The semantics on the other hand is defined by computing a semantic similarity score using lexical database WordNet.

Additionally, this chapter investigates into how much each of contextual and semantic components contributes to the overall task of enriching the domain ontology concepts. Obtained results are compared with $tf*idf$, χ^2 , and LSA. Results for several domains including Computer, Software Engineering, C++ Programming, Database, and the Internet are presented in this chapter.

The rest of the chapter is organized as follows. *Background* section presents ontology enrichment pipeline, describes various input data modalities, discusses text-based resources followed by concept learning techniques and applications using them. This section also presents the state-of-the-art systems in the field of ontology enrichment. Section *SEMCON* describes the proposed SEMCON model in detail. In Section, *Experimental Procedures* we describe the experiments including subjective and objective evaluation of SEMCON along with measures used to evaluate the effectiveness of objective methods.

33 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-hybrid-concept-learning-approach-to-ontology-enrichment/196436

Related Content

Digital Copyright Management Mechanism Based on Dynamic Encryption for Multiplatform Browsers

Ming-Te Chen, Yi Yang Chang and Ta Jen Wu (2024). *International Journal on Semantic Web and Information Systems* (pp. 1-22).

www.irma-international.org/article/digital-copyright-management-mechanism-based-on-dynamic-encryption-for-multiplatform-browsers/334591

Bringing Semantic Services to Real-World Objects

Massimo Paolucci, Gregor Broll, John Hamard, Enrico Rukzio, Matthew Wagner and Albrecht Schmidt (2008). *International Journal on Semantic Web and Information Systems* (pp. 35-49).

www.irma-international.org/article/bringing-semantic-services-real-world/2845

Enabling RDF Stream Processing for Sensor Data Management in the Environmental Domain

Alejandro Llaves, Oscar Corcho, Peter Taylor and Kerry Taylor (2016). *International Journal on Semantic Web and Information Systems* (pp. 1-21).

www.irma-international.org/article/enabling-rdf-stream-processing-for-sensor-data-management-in-the-environmental-domain/164482

An Architecture for Managing Knowledge and System Dynamism in the Worldwide Sensor Web

Deshendran Moodley, Ingo Simonis and Jules Raymond Tapamo (2012). *International Journal on Semantic Web and Information Systems* (pp. 64-88).

www.irma-international.org/article/architecture-managing-knowledge-system-dynamism/70587

Forum Summarization to Support Tutor and Teacher in group interaction management

Antonella Carbonaro (2011). *Semantic Web Personalization and Context Awareness: Management of Personal Identities and Social Networking* (pp. 22-31).

www.irma-international.org/chapter/forum-summarization-support-tutor-teacher/52863