Ontology Enrichment with Causation Relations

Amaal Saleh Hassan Al Hashimy
Department of Computer Science
Sultan Qaboos University
Muscat, Oamn
amaalh@squ.edu.om

Narayanan Kulathuramaiyer
Faculty of Computer Science and Information Technology
University of Malaya Sarawak UNIMAS
Sarawak, Malaysia
naraa@fit.unimas.my

Abstract—Ontology learning is considered a potential approach that can help to reduce the bottleneck of knowledge acquisition. However it suffers from a lack of standards to define concepts, besides the lack of fully automatic knowledge acquisition methods. In performing this learning process, the discovery of non-taxonomic relationships has been identified as being the most difficult. This study is then an attempt to create an enhanced framework for discovering and classifying ontological relationships by using a machine learning strategy. We take into consideration the context of the input text in performing the classification of the semantic relations, in particular, causation relations. The proposed framework extracts initial semantic patterns for causation relation from the input samples, then filters these patterns using two novel algorithms, namely, the "Purpose Based Word Sense Disambiguation" which helps in determining the causation senses for input pair of words and the "Graph Based Semantics" which determines the existence of the causation relations in the sentence and to extract their cause-effect parts. The results show a good performance and the implemented framework cut off many steps of the usual process to produce the final results.

Index Terms— Ontology learning, causation relation, word sense disambiguation.

I. INTRODUCTION

Ontologies have shown their usefulness in application areas such as intelligent information integration, and natural-language processing among others. However, their wide-spread usage is still hindered by ontology acquisition being rather time consuming and, hence, expensive. This is what is known as knowledge bottleneck. For this reason ontology learning from texts constitutes a promising means for ontology acquisition to significantly speed up the ontology building process. As such, several approaches have been proposed for covering the different phases it involves. In this process, the phase of extraction of non-taxonomic relationships (like cause_effect, part_whole) has been recognized as one of the most difficult and least tackled problems (Sánchez & Moreno, 2008).

Even the methods that address non-taxonomic relations did not come up with a widely accepted way in enhancing the process of classifying semantic relations. Most of the systems mainly concentrate on providing methods for creating the relations, but do not consider the context in which the relations might occur in (Hendrickx, 2009).

Non-taxonomic relations between concepts have two main challenges
- Discovering the existence of a relationship between a pair of concepts.
- Labeling this relationship according to its semantic meaning.

The assignment of labels to relationships is difficult since various relationships among instances of the same general concepts are possible. Moreover, even if the semantic is clear, it might still be hard to guess which among several synonymous labels are preferred by a certain community (Kavalec & Svátek, 2005).

According to these facts, this field would greatly benefit if the learning algorithms used to classify the lexical semantic relations, are able to generate new relations automatically based on some learned rules or criteria.

Another source of challenge in this area is that most of the projects have been focused on the construction of the lexical semantic relations within a specific domain and specific components, rather than generating the patterns for these relations (Kim, 2006).

We need a way to detect the semantic patterns that encode a certain relation pattern in text. And, since many of these patterns are ambiguous as they encode more than one relation, we need a way to categorize the relations from these pattern and return only the correct one automatically.

In this research we will focus on the learning of semantic relations patterns between word meanings by taking into consideration the surrounding context in the general domain.

The input resources used for the experimentation include the SemCore, GlossWN, SemEval07, and SemEval10 corpus. The prototype system developed has shown to work without any manual operations to produce general domain causation relations without a need for relation specific knowledge.

II. LITERATURE REVIEW

Causality is a complex, non-primitive relation, which can be refined into more specialized sub-relations. Nastase et al. (2006) classified the causation relation into three kinds as shown in (Figure 1).