Chapter 22

Extracting Commonsense Knowledge Using Concepts Properties

Eduardo Blanco  
The University of Texas at Dallas, USA

Hakki C. Cankaya  
Izmir University of Economics, Turkey

Dan Moldovan  
The University of Texas at Dallas, USA

ABSTRACT

Commonsense knowledge encompasses facts that people know but do not communicate most of the time. For example, one needs water and soap to take a shower is commonsense. This chapter presents a semantically grounded method for extracting commonsense knowledge. First, commonsense rules are identified, e.g., one cannot see imaginary objects. Second, those rules are combined with a basic semantic representation in order to infer commonsense facts, e.g., one cannot see a flying carpet. Further combinations of semantic relations with inferred commonsense facts are proposed and analyzed. Experimental results show that this novel method is able to extract thousands of commonsense facts with little human interaction and high accuracy.

INTRODUCTION

Commonsense knowledge encompasses facts that people know and use in their daily lives. It is assumed to be known by average people; therefore it is not verbally communicated most of the time. For example, when John says I’m going to brush my teeth, he is implicitly saying that most probably he will do that in the bathroom by the sink, will use a toothbrush and toothpaste, will not swallow but spit the toothpaste, will rinse his mouth, and so forth.

It is widely accepted that in order to make machines more intelligent they need to be aware of the vast amount of knowledge that humans have from their early experiences. If machines are to
interact with humans in an intelligent way, they need to have commonsense knowledge (Minsky, 2000). Armed with commonsense knowledge, in the context of the sentence above, a machine would know where John is, what utensils he is using, how long the activity takes, etc.

Commonsense knowledge is defeasible and context dependent, which complicates matters greatly. For example, one can see through windows is commonsense. However, very dirty or dark tinted windows do not allow one to see through. Cameras do not allow to see through clothes, but x-ray cameras do. A sailboat could return to port with a broken engine as long as there is wind. Black clothes are appropriate for a funeral holds in most Western cultures, but the opposite color, white, is appropriate according to Japanese tradition. If German students knock on their desks after a lecture, it was a good lecture, but the same action would be interpreted differently in other countries.

Commonsense knowledge is needed in numerous applications and its unavailability often hinders system performance. For example, applications requiring text understanding and inferences, like question answering, recognizing textual entailments, or extracting implicatures would benefit from commonsense knowledge. Consider the sentence John is cooking dinner. A question answering system would be able to answer the question Where is John likely to be? if it has access to the commonsense fact cooking usually occurs in a kitchen. In Artificial Intelligence, qualitative reasoning, the ability to reason without precise quantitative information (Iwasaki, 1997), and analogical reasoning, the ability to solve problems based on past cases (Rissland, 2006), are but a few examples of commonsense knowledge applications.

In this Chapter, we present a novel method to extract commonsense knowledge by exploiting existing resources. First, we identify commonsense rules (e.g., edible concepts can be found in supermarkets, stores, markets, fridges, plates, etc.) from three sources: human beings, WordNet (Miller, 1995) and ConceptNet (Havasi, Speer, & Alonso, 2007). In our experiments, 98% of rules are automatically extracted from WordNet or ConceptNet. Second, we combine these rules with a basic semantic representation of concepts (e.g., edible is a property of vegetables) in order to obtain commonsense facts (e.g., vegetables can be found in supermarkets). Third, we define two extensions to compose extracted commonsense facts with semantic relations in order to obtain more commonsense facts (e.g., vegetables can be found in a kitchen, lentils are a kind of vegetable; therefore, lentils can be found in supermarkets). The result is thousands of commonsense facts highly accurate.

BACKGROUND

Commonsense knowledge is considered obvious and many researchers (Singh, 2002; Lenat, 1995; Ahn, Kedia, & Blum, 2006) claim that it is unfeasible to obtain it from text or any other existing resource. They believe that humans, experts or non-experts, are needed to obtain commonsense facts.

Cyc is the biggest and oldest project aiming at building a commonsense knowledge base (Lenat, 1995). The project started in 1984 and since then experts have introduced millions of commonsense facts using a formal language, CycL. Currently, the Cyc knowledge base contains nearly 500,000 terms, including about 15,000 types of relations, and about 5,000,000 facts (assertions) relating these terms (Source: http://www.cyc.com/).

ConceptNet is a semantic network for commonsense knowledge generated from the data collected by the Open Mind Common Sense Project (OMCSP) (Havasi, Speer, & Alonso, 2007). It started in 2000 and it is based on a collaborative effort of thousands of anonymous non-expert users over the internet. The OMCSP project has collected over 700,000 pieces of common sense