Using Semantic Information to Improve Case Retrieval in Case-Based Reasoning Systems

J. Akshay Iyer and Pushpak Bhattacharyya

Indian Institute of Technology Bombay
{akshay,pb}@cse.iitb.ac.in

Abstract. Conventional Case-Based Reasoning (CBR) systems rely on word knowledge to index and search cases from its memory. On being presented with a problem, the Case-Based Reasoning system tries to retrieve a relevant case based on the words that appear in the problem sentence without considering their respective senses. Drawbacks of such systems become more evident in cases where the input is in the form of a sentence in a natural language. Ignoring semantic information in this case may not result in retrieval of desired case or may result in retrieval of an undesired case. In this paper we present a method that tries to improve the precision of retrieval by also taking into account semantic information available to us about the words in the problem sentence. Towards this goal, Universal Networking Language (UNL) is made use of, which provides a semantic representation of natural language text to capture sentence structure. Lexical resource like WordNet is used for finding semantic similarity between two concepts. Using an existing commercial Case-Based Reasoning system as basis for comparison, we demonstrate that considering such semantic information helps in improving case retrieval.

1 Introduction

Case-Based Reasoning (CBR) Systems are one of the most widely used systems in the field of problem solving and planning. A number of such systems are developed and reported [5, 8, 11]. Typically, a number of cases are stored in memory and upon being presented with a problem, a set of relevant cases is retrieved and presented as a solution to the problem [7]. One of the fundamental issues in such systems concerns this retrieval process. Information from the input problem is extracted out and this information is used to index (or search) in the memory to locate the desired case. In systems where a problem is input in Natural Language form, the issue becomes more profound. Traditionally, a number of statistical methods are used for extracting information from the input problem and using it in turn for identifying cases that are relevant to the problem. However, since such methods do not employ any natural language understanding, they fail in situations when mere knowledge about words is not sufficient.

In this paper we propose a method by which we could use information, both semantic and syntactic, from natural language text to compare and retrieve relevant cases. The rest of the paper is organized as follows. Section 2 discusses the shortcom-