Poster presented at the 5^{th} Annual IRIS/PRECARN Conference, Toronto, ON, June 13-15, 1995

CASE-BASED REASONING SYSTEM APPLIED AS AN INTELLIGENT RETRIEVAL AND ADVISORY TOOL

Igor Jurišica, PhD Degree Student
Department of Computer Science
University of Toronto, Toronto, Ontario M5S 1A4

Intelligent information retrieval helps to find relevant information based on imprecisely specified query. However, finding a relevant piece of knowledge is only a first step in using it. Case-based reasoning systems can be used to solve new problems based on retrieved information similarly as human experts do – if the current problem is similar to a solved case, then its solution can be adapted and reused to solve a new problem.

For retrieval of relevant information we designed a novel context-based similarity retrieval algorithm. Context specifies what part of an information to consider and what matching criteria to use. Context also defines relevance of the retrieved items. It can be used to control quality (relevance) and quantity of retrieved information.

The results of system evaluation on IVF domain suggest that context-based retrieval is effective in case-based retrieval and diagnosis. After knowing only 8 attribute/value pairs, the system predicts values for the remaining attributes. Successful cases were used to suggest treatment; unsuccessful cases served as a warning. The accuracy of prediction can be used as measure of the retrieval effectiveness of the system – relevant cases provide for accurate predictions. Using a simple cross-validation (leave-one-out method) and averaging results over 20 random trials, a relative error that falls in the range of 0.15 +/- 0.09 and 0.07 +/- 0.04 with 95performance could be improved if the database contained additional important data needed for problem solving (if its informativeness was increased).

Evaluation of the system on a robotic domain was targeted on prediction. Retrieved relevant items were used to predict a real value for an attribute with 95error 0.056 +/- 0.009 was observed. The result compares favorably to other machine learning algorithm applied to the same problem.

An evaluation of context-based retrieval was also performed on a small software repository. Performance evaluation in this domain is more subjective due to the subjectivity of relevance measure. This experiment demonstrated, however, how the added flexibility of the proposed system could be utilized to alter the system's performance from precision to recall-oriented, i.e., from requiring that no irrelevant cases are retrieved (system is allowed to miss some relevant cases) through to requiring that absolutely all relevant cases are retrieved (system is allowed to retrieve some irrelevant cases).

From the application of context-based retrieval to a number of datasets, we have concluded that if the user is an expert in the domain (he/she knows salience and value ranges for attributes), then value relaxation is the most suitable form of relaxation to apply in order to retrieve additional cases. Value relaxation allows for increasing the precision while preserving perfect recall. If, however, the user is not as confident, then an iterative process involving cardinality relaxation may be the most effective way to retrieve relevant items. In this situation the system retrieves relevant items with perfect recall, but decreased precision.