Chapter VII
Flexible Querying Techniques Based on CBR

Guy de Tré
Ghent University, Belgium

Marysa Demoor
Ghent University, Belgium

Bert Callens
Ghent University, Belgium

Lise Gosseye
Ghent University, Belgium

ABSTRACT

In case-based reasoning (CBR), a new untreated case is compared to cases that have been treated earlier, after which data from the similar cases (if found) are used to predict the corresponding unknown data values for the new case. Because case comparisons will seldom result in an exact-similarity matching of cases and the conventional CBR approaches do not efficiently deal with such imperfections, more advanced approaches that adequately cope with these imperfections can help to enhance CBR. Moreover, CBR in its turn can be used to enhance flexible querying. In this chapter, we describe how fuzzy set theory can be used to model a gradation in similarity of the cases and how the inevitable uncertainty that occurs when predictions are made can be handled using possibility theory resulting in what we call flexible CBR. Furthermore, we present how and under which conditions flexible CBR can be used to enhance flexible querying of regular databases.

INTRODUCTION

In case-based reasoning (CBR), knowledge is deduced from the characteristics of a collection of past cases, rather than induced from a set of knowledge rules that are stored in a knowledge base. In this way, CBR can be applied to find the solution of a given problem on the basis of the known solutions to similar problems. Of course, this can only be done if it holds that:

‘Similar problems have similar solutions’
This statement is the underlying hypothesis of CBR (Aamodt & Plaza, 1994). Without this hypothesis, CBR cannot be used. Starting from this hypothesis, the problem of finding the solution or outcome of a new case is solved by matching the characteristics of this case against those of similar cases. The result of this matching process then allows predicting the solution or outcome. At a later stage, when more information becomes available, the predictions could be tested as to their correctness, and if necessary, the process to find similar cases could be revised. The hypothesis thus allows solving problems without necessitating the explicit modeling of expert knowledge because in CBR this knowledge is implicit in the solution.

CBR techniques have been successfully applied in many application fields. Under specific circumstances, they can also be applied to enhance and enrich database querying (de Calmès, Dubois, Hübnermeier, Prade, & Sedes, 2003; Ellman, 1995; Shimazu, Kitano, & Shibata, 1993). This chapter deals with the application of CBR in order to enhance the querying and accessibility of regular databases. A precondition for applying CBR techniques in database querying is that the database must contain comparable descriptions of real cases that all relate to the same topic and have common characteristics. In the remainder of the chapter, it is, without a loss of generality, assumed that each case description consists of attribute values that each describe a characteristic of the case. For new cases, some of the attribute values might be unknown. This can, for example, be due to the fact that the case was not completely handled/described at the time when it was initially stored in the database. If the underlying hypothesis of CBR holds for the problem of value prediction for an attribute, CBR techniques can also be used to predict the unknown values for that attribute.

This capability to predict unknown attribute values can be fully exploited to enhance flexible querying of both regular and fuzzy databases. If users are looking for case descriptions for which an attribute takes a given value or has a value that is within a given range of values, then value prediction also allows finding those cases which might have the requested value in the future. Of course, this kind of flexibility requires an enriched querying mechanism that allows the modeling of uncertain query results (due to the fact that it is obtained from a prediction).

Because similarity between two cases is rarely a matter of all or nothing, but rather a matter of degree, such an enriched CBR mechanism or CBR based querying mechanism should also support the modeling of imprecision stemming from case comparison. Fuzzy set theory (Dubois & Prade, 2000; Pedrycz & Gomide, 1998; Zadeh, 1965) can be used to model such kinds of imprecision. This is especially the case because there is a close connection between fuzzy-set based approximate reasoning and the underlying inference principle of CBR (Dubois, Esteva, Garcia, Godo, Lopez de Mantaras, & Prade, 1998; Yager, 1997). Moreover, using fuzzy set theory also has the advantage that the related possibility theory (Dubois & Prade, 1988; Zadeh, 1978) can be used to model the uncertainty that is inherent to prediction (Dubois, Hübnermeier, & Prade, 2000).

In this chapter, we describe an enhanced CBR-based approach for flexible querying of regular databases. The approach is based on fuzzy set theory and possibility theory and enables the prediction of unknown attribute values of case descriptions that are inserted in a database system on condition that the underlying hypothesis of CBR holds in the context of the prediction problem. As an example, consider a regular relational database in which information about juridical complaints, as registered by lawyers after interaction with the aggrieved party is stored. In a juridical context, it holds that ‘similar complaints must be dealt with in a similar way’ and thus, by consequence, the underlying hypothesis of CBR also holds.

The user has to select the ‘descriptive’ attributes on which the case comparison will be based. In the complaint database, the ‘descriptive’ attributes could be the attributes that are used to classify the complaint, the age of the victim, the gender of the victim, the address of the victim, the job of the vic-