A Probabilistic Exemplar-Based Model for Case-Based Reasoning

Andrés F. Rodríguez¹, Sunil Vadera², and L. Enrique Sucar³

¹ Instituto de Investigaciones Eléctricas
AP 475, Cuernavaca, Mor. México
afrm@iie.org.mx

² School of Sciences, University of Salford
Salford, M5 4WT, UK
S.Vadera@cms.salford.ac.uk

³ Departamento de Computación, ITESM - Campus Morelos
AP C-99, Cuernavaca, Mor., 62020, México
esucar@campus.mor.itesm.mx

Abstract. An exemplar-based model with foundations in Bayesian networks is described. The proposed model utilises two Bayesian networks: one for indexing of categories, and another for identifying exemplars within categories. Learning is incrementally conducted each time a new case is classified. The representation structure dynamically changes each time a new case is classified and a prototypicality function is used as a basis for selecting suitable exemplars. The results of evaluating the model on three datasets are presented.

1 Introduction

Organizing and indexing cases in memory is a fundamental part of case-based reasoning (CBR) that involves learning and reasoning processes. This problem can be divided into two parts. The first is the selection of the features of the cases that can be used to index and retrieve the cases. The second is the organisation of the case memory so that the case retrieval process is efficient and accurate.

One approach that has been used to address this problem has been to store all the cases and develop algorithms to partition the search space for retrieving similar cases. So for example, systems like REMIND provide a tree induction algorithm that can be used to avoid examining all the cases. This kind of approach is particularly useful when large databases of cases are already available. However, when cases are not available in advance, and the domain is not well defined this approach is more difficult to apply.

An alternative approach, that is perhaps more applicable to such situations, is to store only prototypical cases. This approach, known as the exemplar-based model has its basis in cognitive theories, which postulate that concepts can be represented by exemplars [13,14]. However, previous implementations of the exemplar-based models have struggled to produce systems that can be justified in a rational manner. For example, the Protos system [3] uses many heuristics...
and mechanisms for combining evidence that are hard to justify independently of its original application.

Hence, this paper attempts to develop an exemplar-based model with foundations that utilise Bayesian models.

2 Exemplar-Based Models: The Problem

An exemplar-based model is thought to be particularly appropriate for weak domains [11], where it is difficult to define categories by explicitly using classical membership constraints. For such situations, which can be common in real applications, an exemplar-based model may provide a better representation of the categories. Figure 1 illustrates the idea of a weak domain together with cases, exemplars, and categories. It shows two categories, A, B (the solid lines), that each have exemplars that represent regions (the dashed lines) that contain cases (the dots).

![Fig. 1. Cases, exemplars and categories in a weak domain.](image)

To illustrate the problem, suppose that the category A is represented by the exemplars $e_1, e_2$, and $e_3$ and the category B is represented by the exemplars $e_3$ and $e_4$. Also suppose that the exemplars $e_1, e_2, e_3$, and $e_4$ currently represent 4, 2, 3, and 2 cases respectively. Now suppose that a new case is given. The following two functions must be provided by an exemplar-based model:

1. Determine the exemplar that best classifies the new case given the available information.
2. Determine how knowing the new instance and its classification it can be used to improve the accuracy of the model.

The first of these functions is a classification task, while the second can be viewed as a supervised learning task [1].