Chapter 1.3
Machine Learning Through Data Mining

Diego Liberati
Italian National Research Council, Italy

INTRODUCTION

In dealing with information it often turns out that one has to face a huge amount of data, often not completely homogeneous and often without an immediate grasp of an underlying simple structure. Many records, each one instantiating many variables, are usually collected with the help of various technologies.

Given the opportunity to have so many data not easy to correlate by the human reader, but probably hiding interesting properties, one of the typical goals one has in mind is to classify subjects on the basis of a hopefully reduced meaningful subset of the measured variables. The complexity of the problem makes it worthwhile to resort to automatic classification procedures.

Then, the question arises of reconstructing a synthetic mathematical model, capturing the most important relations between variables, in order to both discriminate classes of subjects and possibly also infer rules of behaviours that could help identify their habits.

Such interrelated aspects will be the focus of the present contribution. The data mining procedures that will be introduced in order to infer properties hidden in the data are in fact so powerful that care should be put in their capability to unveil regularities that the owner of the data would not want to let the processing tool discover, like for instance, in some cases the customer habits investigated via

DOI: 10.4018/978-1-60960-818-7.ch1.3

the usual smart card used in commerce with the apparent reward of discounting.

Four main general purpose approaches will be briefly discussed in the present article, underlying the cost effectiveness of each one.

In order to reduce the dimensionality of the problem, simplifying both the computation and the subsequent understanding of the solution, the critical issues of selecting the most salient variables must be addressed. This step may already be sensitive, pointing to the very core of the information to look at.

A very simple approach is to resort to cascading a divisive partitioning of data orthogonal to the principal directions (PDDP) (Boley, 1998) already proven to be successful in the context of analyzing micro-arrays data (Garatti, Bittanti, Liberati, & Maffezzoli, 2007).

A more sophisticated possible approach is to resort to a rule induction method, like the one described in Muselli and Liberati (2000). Such a strategy also offers the advantage to extract underlying rules, implying conjunctions or disjunctions between the identified salient variables. Thus, a first idea of their even nonlinear relations is provided as a first step to design a representative model, whose variables will be the selected ones. Such an approach has been shown (Muselli & Liberati, 2002) to be not less powerful over several benchmarks than the popular decision tree developed by Quinlan (1994). An alternative in this sense can be represented by Adaptive Bayesian networks (Yarmus, 2003) whose advantage is also to be available on a commercial wide spread data base tool like Oracle.

Dynamics may matter. A possible approach to blindly build a simple linear approximating model is thus to resort to piece-wise affine (PWA) identification (Ferrari-Trecate, Muselli, Liberati, & Morari, 2003).

The joint use of (some of) such four approaches briefly described in this article, starting from data without known priors about their relationships, will allow to reduce dimensionality without significant loss in information, then to infer logical relationships, and, finally, to identify a simple input-output model of the involved process that also could be used for controlling purposes, even those potentially sensitive to ethical and security issues.

BACKGROUND

The introduced tasks of selecting salient variables, identifying their relationships from data, and classifying possible intruders may be sequentially accomplished with various degrees of success in a variety of ways:

- Principal components order the variables from the most salient to the least one, but only under a linear framework.
- Partial least squares do allow to extend to nonlinear models, provided that one has prior information on the structure of the involved nonlinearity; in fact, the regression equation needs to be written before identifying its parameters.
- Clustering may operate even in an unsupervised way without the a priori correct classification of a training set (Boley, 1998).
- Neural networks are known to learn the embedded rules with the indirect possibility (Taha & Ghosh, 1999) to make rules explicit or to underline the salient variables.
- Decision trees (Quinlan, 1994) are a popular framework providing a satisfactory answer to the recalled needs.

RECENT DEVELOPMENTS

Unsupervised Clustering

In this contribution, we will firstly resort to a quite recently developed unsupervised clustering approach, the Principal Direction Divisive Partition-