Rule Qualities and Knowledge Combination for Decision-Making

Ivan Bruha
McMaster University, Canada

INTRODUCTION

Within the past several years, research in decision-supporting systems has been investigating the possibilities of enhancing their overall performance, particularly their prediction (classification) accuracy, or performance quality, and their time complexity. One such discipline, data mining (DM), processes usually very large databases in a profound and robust way. Since data are collected and stored at a very large acceleration these days, there has been an urgent need for a new generation of robust software packages to extract useful information or knowledge from large volumes of data. Research is supposed to develop methods and techniques to process large data in order to receive knowledge, which is hidden in these databases, that would be compact, more or less abstract, but understandable and useful for further applications.

DM usually is defined as a nontrivial process of identifying valid, novel, and ultimately understandable knowledge in data. It is understood that DM points to the overall process of determining a useful knowledge from databases (i.e., extracting high-level knowledge from low-level data in the context of large databases). It can be viewed as a multi-disciplinary activity, because it exploits several research disciplines of artificial intelligence (AI), such as machine learning, pattern recognition, expert systems, and knowledge acquisition, as well as mathematical disciplines such as statistics, theory of information, and uncertainty processing.

This article discusses two enhancements in DM: rule quality and knowledge integration/combination in the section, Main Thrust of the Article. The future possible directions in these two fields are briefly discussed in the next to the Future Trends section. The last section then analyzes the enhancements achieved by embedding the measures into rule-based classifiers and the multi-strategy approach in decision-supporting systems.

It also should be noted that there is no uniform terminology in the knowledge-intensive systems (including DM and machine learning, of course); therefore, here, we usually use not a single term, but several most common terms that can be found in literature. Also, some definitions are not uniform but overlap (see the section, Terms and Definitions).

BACKGROUND

Data Mining (DM) or Knowledge Discovery in Databases (KDD) utilizes several paradigms for extracting a knowledge that then can be exploited as a decision scenario (architecture) within expert or classification (prediction) systems. One commonly used paradigm in Machine Learning (ML) is called *divide and conquer*, which induces decision trees. Another widely used *covering paradigm* generates sets of decision rules (e.g., the CNx family [Clark & Niblett, 1989], C4.5Rules, Ripper, etc.). However, the rule-based classification systems are faced by an important deficiency that is to be solved in order to improve the predictive power of such systems.

The traditional decision-making systems have been dependent on a single technique, strategy, or architecture. Therefore, their accuracy and successfulness have not been so high. New sophisticated decision-supporting systems utilize results obtained from several lower-level systems, each usually (but not required to be) based on a different paradigm, or combine or refine them within a dynamic process. Thus, such a multi-strategy (hybrid) system consists of two or more individual agents that interchange information and cooperate together.

It should be noted that there are, in fact, two fundamental approaches for combining the information from multi-data tasks:

1. In data combination, the datasets are merged into a single set before the actual knowledge acquisition.
2. In knowledge (theory) combination, or sensor fusion, several agents (base classifiers, sensors) process each input dataset separately, and the induced models (knowledge bases) then are combined at the higher level.

The next section discusses the latter approach, including the more general aspect of knowledge integration. There are various knowledge combination schemes (e.g., the best, weighted voting, sensitive voting, Bayesian combination, etc.). The next section focuses on relatively new trends in knowledge combination.

Furthermore, there are two types of agents in the multi-strategy (knowledge combination) decision-supporting

architecture. The simpler one yields a single decision; the more sophisticated one induces a list of several decisions. In both types, each decision should be accompanied by the agent's confidence (belief) in it. These functional measurements are supported mostly by statistical analysis that is based on both the certainty (accuracy, predictability) of the agent as well as the consistency of its decision. There have been quite a few research inquiries to define formally such statistics; some, however, have yielded in quite complex and hardly enumerable formulas, so that they have never been used. The following section presents a simpler but more understandable approach to define these measurements.

MAIN THRUST AND BACKGROUND

(a) Rule Quality

A rule-inducing algorithm may yield either an ordered or unordered set of decision rules. The latter seems to be more understandable by humans and directly applicable in most decision-supporting systems. However, the classification utilizing an unordered set of decision rules exhibits a significant deficiency, not immediately apparent. Three cases are possible:

1. If an input unseen (to-be-classified) object satisfies (matches, fires for) one or more rules of the same class, then the object is categorized to the class assigned to the rule(s).
2. If the unseen object is not covered by any rule, then either the classifier informs the user about its inability to decide ('I do not know'), or the object is assigned by default to the majority class in the training set, or some similar techniques are invoked.
3. Difficulty arises if the input object satisfies more rules assigned to different classes. Then, some schemes have to be applied to assign the unseen input object to the most appropriate class.

One possibility to clarify the conflict situation (case 3) of multiple-rule systems is to associate each rule in the decision scheme (knowledge base, model) of a classifier with a numerical factor that can express its properties and characterize a measure of belief in the rule, its power, predictability, reliability, likelihood, and so forth. A collection of these properties is symbolized by a function commonly called the *rule quality*. After choosing a formula for the rule quality, we also have to select a scheme for combining these qualities (quality combination).

Quality of rules, its methodology, as well as appropriate formulas have been discussed for many years. Bergadano, et al. (1992) is one the first papers that intro-

duces various definitions and formulas for the rule quality; besides the rule's power and predictability, it measures its size, understandability, and the like. Formulas for the rule quality have been studied and tested further in several other papers (An & Cercone, 2001; Hipp et al., 2002). A survey of the rule combinations can be found in Kohavi and Kunz (1997).

Comprehensive analysis and empirical expertise of formulas of rule qualities and their combining schemes have been published in Bruha and Tkadlec (2003) and their theoretical methodology in Tkadlec and Bruha (2003). The first one introduces quite a few statistical and empirical formulas for the rule quality, including the quality combinations, and compares them. A rule quality, in most cases, is a function of its consistency (sensitivity), completeness (coverage, positive predictive value), and other statistics, such as a rule's matching rates. Because we deal with real-world noisy data, any decision set induced must be not only reliable but also powerful. Its reliability is characterized by a consistency factor and its power by a completeness. These and other statistical factors usually are defined by means of the so-called 2×2 contingency table.

The latter paper introduces theoretical formalism and methodological tools for building multiple-rule systems. It focuses on four agents that cooperate with each other: designer, learner, classifier, and predictor. The paper offers to a designer of a new multiple-rule system the minimum requirements for the previously discussed concepts and (mostly statistical) characteristics that the designer can start with. It also exhibits a general flow chart for a decision-system builder.

In addition to the rule quality discussed previously, there are other rule measures, such as its size (i.e., the number of attribute pairs involved), computational complexity, comprehensibility ('Is the rule telling humans something interesting about the application domain?'), understandability, redundancy (measured within the entire decision set of rules), and the like (Tan, Kumar & Srivastava, 2004). However, some of these characteristics are subjective; on the contrary, formulas for rule quality are supported by theoretical sources or profound empirical expertise.

In most decision-supporting systems, the rule qualities are static, constant, and calculated a priori before the actual classification or prediction. Their predictability can be improved by a dynamic change of their values during the classification process. One possible scheme implants a feedback loop from the classifier to the learner (Bruha, 2000); it refines (modifies) the rule qualities according to the correct/false predictions made by the classifier by changing the qualities of the rules that were involved in the current classification. The entire refinement method thus may be viewed as a (semi-) meta-

Related Content

Data Quality in Cooperative Information Systems

Carla Marchetti, Massimo Mecella, Monica Scannapiecoand Antoninio Virgillito (2005). *Encyclopedia of Data Warehousing and Mining (pp. 297-301).*

www.irma-international.org/chapter/data-quality-cooperative-information-systems/10611

Warehousing and Mining Streams of Mobile Object Observations

S. Orlando, A. Raffaetà, A. Roncatoand C. Silvestri (2010). *Intelligent Techniques for Warehousing and Mining Sensor Network Data (pp. 72-98).*

www.irma-international.org/chapter/warehousing-mining-streams-mobile-object/39541

Data Mining and Decision Support for Business and Science

Auroop R. Ganguly, Amar Guptaand Shiraj Khan (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2618-2625).*

www.irma-international.org/chapter/data-mining-decision-support-business/7786

Mining Images for Structure

Terry Caelli (2005). *Encyclopedia of Data Warehousing and Mining (pp. 805-809).*

www.irma-international.org/chapter/mining-images-structure/10707

Materialized Hypertext View Maintenance

Giuseppe Sindoni (2005). *Encyclopedia of Data Warehousing and Mining (pp. 710-713).*

www.irma-international.org/chapter/materialized-hypertext-view-maintenance/10689