Information Sciences 178 (2008) 4019–4037

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Stochastic dominance-based rough set model for ordinal classification

Wojciech Kotłowski [a], Krzysztof Dembczyński [a], Salvatore Greco [b], Roman Słowiński [a,c,*]

[a] Institute of Computing Science, Poznań University of Technology, Piotrowo 2, Poznań 60-965, Poland
[b] Faculty of Economics, University of Catania, 95129 Catania, Italy
[c] Institute for Systems Research, Polish Academy of Sciences, Warsaw 01-447, Poland

ARTICLE INFO

Article history:
Received 12 November 2007
Received in revised form 18 April 2008
Accepted 7 June 2008

Keywords:
Dominance-based rough set approach
Ordinal classification
Monotonicity constraints
Isotonic regression
Maximum likelihood estimation
Variable consistency models
Statistical decision theory
Empirical risk minimization
Multiple criteria decision analysis

ABSTRACT

In order to discover interesting patterns and dependencies in data, an approach based on rough set theory can be used. In particular, dominance-based rough set approach (DRSA) has been introduced to deal with the problem of ordinal classification with monotonicity constraints (also referred to as multicriteria classification in decision analysis). However, in real-life problems, in the presence of noise, the notions of rough approximations were found to be excessively restrictive. In this paper, we introduce a probabilistic model for ordinal classification problems with monotonicity constraints. Then, we generalize the notion of lower approximations to the stochastic case. We estimate the probabilities with the maximum likelihood method which leads to the isotonic regression problem for a two-class (binary) case. The approach is easily generalized to a multi-class case. Finally, we show the equivalence of the variable consistency rough sets to the specific empirical risk-minimizing decision rule in the statistical decision theory.

1. Introduction

We consider an *ordinal classification problem* that consists in assignment of objects to K ordered *classes* Cl_k, $k \in Y = \{1, \ldots, K\}$, such that if $k > k'$ then class Cl_k is higher than class $Cl_{k'}$. Objects are evaluated on a set of m *attributes* with ordered value sets. Here, without loss of generality, we assume that the value set of each attribute is a subset of $\mathbb{R}$ (even if the scale is purely ordinal, evaluation on attributes can be numbercoded) and the order relation is a linear order $\geqslant$, so that each object x_i is an m-dimensional vector $(x_{i1}, \ldots, x_{im})$. It is assumed that *monotonicity constraints* are present in the data: a higher evaluation of an object on an attribute, with other evaluations being fixed, should not decrease its assignment to the class. One can induce a data model from a *training* set $U = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, consisting of n objects (denoted with x) already assigned to their classes (class indices denoted with $y \in Y$). We also denote $X = \{x_1, \ldots, x_n\}$, and by class Cl_k we mean the subset of X consisting of objects x_i having class indices $y_i = k$, $Cl_k = \{x_i \in X : y_i = k\}$.

Thus, ordinal classification problem with monotonicity constraints resembles a typical classification problem considered in machine learning [10,17], but requires two additional constraints. The first one is the assumption of the ordinal scale on each attribute and on class indices. The second constraint is the monotonicity property: the expected class index increases with increasing evaluations on attributes. Such properties are commonly encountered in real-life applications, yet rarely taken into account. In decision theory, a *multicriteria classification problem* is considered [13], which has exactly the form

* Corresponding author. Address: Institute of Computing Science, Poznań University of Technology, Piotrowo 2, Poznań 60-965, Poland.
E-mail addresses: wkotlowski@cs.put.poznan.pl (W. Kotłowski), kdembczynski@cs.put.poznan.pl (K. Dembczyński), salgreco@unict.it (S. Greco), rslowinski@cs.put.poznan.pl (R. Słowiński).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2008.06.013

of ordinal classification problem with monotonicity constraints. Moreover, in many different domains monotone properties follow from the domain knowledge about the problem and should not be neglected. They have been recognized in applications such as bankruptcy risk prediction [11], breast cancer diagnosis [25], house pricing [23], credit rating [9], liver disorder diagnosis [26] and many others.

As an example, consider the customer satisfaction analysis [15], which aims at determining customer preferences in order to optimize decisions about strategies for launching new products, or about improving the image of existing products. The monotonicity constraints are of fundamental importance here. Indeed, consider two customers, *A* and *B*, and suppose that the evaluations of a product by customer *A* on a set of attributes are better than the evaluations by customer *B*. In this case, it is reasonable to expect that also the comprehensive evaluation of this product (i.e. class, to which the product is assigned) by customer *A* is better (or at least not worse) than the comprehensive evaluation made by customer *B*. As another example, consider the problem of credit rating. One of the attributes could be the degree of regularity in paying previous debts by a consumer (with ordered value set, e.g. "unstable", "acceptable", "very stable"); on the other hand, the class attribute could be the evaluation of potential risk of lending money to a consumer, also with ordered value set (e.g. "high-risk", "medium-risk", "low-risk"); moreover, there exists a natural monotone relationship between the two attributes: the more stable the payment of the debt, the less risky the new credit is.

Despite the monotone nature of the data, it still may happen that in the training set U, there exists an object x_i not worse than another object x_j on all attributes, however, x_i is assigned to a class worse than x_j; such situation violates the monotone properties of the data, so we shall call objects x_i and x_j *inconsistent*. Rough set theory [19,20,22] has been adapted to deal with this kind of inconsistency and the resulting methodology has been called *dominance-based rough set approach* (DRSA) [12,13]. In DRSA, the classical indiscernibility relation has been replaced by a dominance relation. Using the rough set approach to the analysis of multicriteria classification problem, we obtain lower and upper (rough) approximations of unions of classes. The difference between upper and lower approximations shows inconsistent objects with respect to the dominance principle. It can happen that due to the presence of noise, the data is so inconsistent, that too much information is lost, thus making the DRSA inference model not accurate. To cope with the problem of excessive inconsistency, a *variable consistency* model within DRSA has been proposed (VC-DRSA) [14].

In this paper, we look at DRSA from a different point of view, identifying its connections with statistics and statistical decision theory. We start with the overview of the classical rough set theory and show that the variable-precision model [31,32] comes from the maximum likelihood estimation method. Then we briefly present main concepts of DRSA. Afterwards, the main part of the paper follows: we introduce the probabilistic model for a general class of ordinal classification problems with monotonicity constraints, and we generalize lower approximations to the stochastic case. Using the maximum likelihood method we show how the probabilities can be estimated in a nonparametric way. It leads to the statistical problem of isotonic regression, which is then solved by the optimal objects reassignment problem. Finally, we explain the approach as being a solution to the problem of finding a decision function minimizing the empirical risk [2].

We stress that the theory presented in this paper is related to the training set only. In order to properly classify objects outside the training set, a generalizing classification function must be constructed. We do not consider this problem here. The aim of this paper is the analysis of inconsistencies in the dataset, handling and correcting them according to the probabilistic model assumption, which comes from exploring the monotonicity constraints. This analysis can be seen as a stochastic extension of DRSA. Therefore, the methodology presented here can be treated as a form of preprocessing and improving the data.

2. Maximum likelihood estimation in the classical variable precision rough set approach

We start with the classical rough set approach [19], which neither takes into account monotonicity constraints nor are the classes and attribute values ordered. It is based on the assumption that objects having the same description are indiscernible (similar) with respect to the available information. The *indiscernibility* relation induces a partition of the universe into blocks of indiscernible objects, called *granules* [19,13]. The indiscernibility relation I is defined as

$$I = \{(x_i, x_j) \in X \times X : x_{it} = x_{jt} \ \forall t = 1, \ldots, m\}, \tag{1}$$

where x_{it} is the evaluation of object x_i on attribute t, as defined in previous section. The equivalence classes of I are called *granules*. The equivalence class for an object $x \in X$ is denoted $I(x)$. Any subset S of the universe may be expressed in terms of the granules either precisely (as a union of granules) or approximately only. In the latter case, the subset S may be characterized by two ordinary sets, called *lower* and *upper approximations*. Here, we always assume, that the approximated set S is a class $Cl_k, k \in Y$. The lower and upper approximations of class Cl_k are defined, respectively, by

$$\underline{Cl}_k = \{x_i \in X : I(x_i) \subseteq Cl_k\}, \tag{2}$$

$$\overline{Cl}_k = \{x_i \in X : I(x_i) \cap Cl_k \neq \emptyset\}. \tag{3}$$

It follows from the definition, that $\underline{Cl}_k$ is the largest union of the granules included in Cl_k, while $\overline{Cl}_k$ is the smallest union of the granules containing Cl_k [19]. It holds, that $\underline{Cl}_k \subseteq Cl_k \subseteq \overline{Cl}_k$. Therefore, if an object $x \in X$ belongs to $\underline{Cl}_k$, it is also certainly an element of Cl_k, while if x belongs to $\overline{Cl}_k$, it may belong to class Cl_k.