Information Sciences 424 (2018) 104–117

Contents lists available at ScienceDirect

Information Sciences

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

Partially monotonic decision trees

Shenglei Pei [a,b], Qinghua Hu [a,*]

[a] School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
[b] School of Physics and Electronic Information Engineering, Qinghai University for Nationalities, Xining 810007, China

ARTICLE INFO

Article history:
Received 8 December 2016
Revised 19 June 2017
Accepted 3 October 2017
Available online 3 October 2017

Keywords:
Decision tree
Partially monotonic
Rank inconsistency
Monotonic directions

ABSTRACT

In multicriteria decision tasks, certain features are linearly ordered according to the decision and are called criteria, whereas others, called regular attributes, are not. In practice, regular attributes and criteria coexist in most classification tasks. In this paper, we propose a rank-inconsistent rate that distinguishes attributes from criteria. Furthermore, it represents the directions of the monotonic relationships between criteria and decisions. We design a partially monotonic decision tree algorithm to extract decision rules for partially monotonic classification tasks. Experimental results show that the proposed algorithm is effective and efficient.

1. Introduction

Monotonic classification is used in many applications, such as multicriteria decision making [3,37,39,40], credit rating [11,14,17], the customer satisfaction analysis [21], and the house pricing problem [42]. There is a wide range of monotonic problems, in which the decision values must increase with the feature values. For instance, the price of a house typically increases with the house's size. Smoking increases the probability of lung cancer. In these classification tasks, monotonic classification functions guarantee monotonicity between attributes and decisions. That is, objects with better feature values should not be assigned worse decision values.

Researchers in monotonic classification have proposed several methods with which to learn and extract decision rules for generating decision models. These studies can be roughly classified into those that involve constructing a theoretical framework with monotonicity constraints, and those that consist of a model-based approach. The dominance-based rough set approach (DRSA) builds a formal framework in which to discuss monotonic classification. This model was proposed by Greco et al. [19,20,22] and applied to monotonic tasks. DRSA was subsequently extensively discussed by other researchers. Błaszczyński et al. [5,6] proposed the variable- consistency dominance-based rough set approach (VC-DRSA) based on extended lower approximation. Soon afterward, they designed a transformation method for discovering partially monotonic relationships. Moreover, the VC-DomLEM algorithm was proposed based on this method [7,8]. This method was also applied to DRSA and VC-DRSA for nonmonotonic decision tasks as well [4]. However, this method is complicated, as it clones attributes with unknown monotonicity. In 2015, Wang et al. [43] proposed a refined method to improve the performance of VC-DomLEM algorithms. The preprocessing method in this approach does not clone all attributes with unknown monotonicity to improve performance. The monotonic k-nearest neighbors (kNN) algorithm was proposed in [15]. It relabels training data and predicts class labels by utilizing modified nearest-neighbor rules. Some hybrid approaches with monotonicity constraints have also been proposed to extract rules in recent years [10,30].

* Corresponding author.
E-mail addresses: peishenglei@qhmu.edu.cn (S. Pei), huqinghua@tju.edu.cn (Q. Hu).

https://doi.org/10.1016/j.ins.2017.10.006
0020-0255/© 2017 Published by Elsevier Inc.

Table 1
An example of overdue evaluation.

Objects	Housing loans	Amount	Usage rates	History overdue	Overdue rank
$\mathbf{x}_1$	Yes	10	60	No	0
$\mathbf{x}_2$	Yes	10	80	No	1
$\mathbf{x}_3$	No	5	100	No	1
$\mathbf{x}_4$	No	5	60	Yes	2
$\mathbf{x}_5$	Yes	1	80	Yes	2

The model-based approach is an alternative for solving classification tasks with monotonicity constraints. Support vector machines (SVMs) and other methods based on kernel learning have recently been applied to monotonic classification problems as well. In 2005, Pelckmans et al. [36] designed monotonic kernel regression based on Least Squares SVM (LS-SVM) regression. This algorithm mainly solves ordinal regression problems. Doumpos et al. [14] proposed a linear SVM with an $L_1 - \infty$ norm for credit risk evaluation, which implied the monotonicity of credit risk. The monotonic SVM model is built by adding monotonicity constraints to the conventional SVM model. An extension of the monotonic SVM was studied in [32]. In [12], monotonic neural networks were proposed for partially monotonic classification. This method contains all available features needed to achieve superior performance over standard neural networks. Decision tree models have been widely applied to monotonic classification tasks. In 1995, Ben-David extended ID3 algorithms to monotonic classification [2]. Since then, monotonic decision trees have been designed to solve monotonic problems of various kinds. A postpruning classification tree was proposed by Feelders [16]. This algorithm prunes the parent node of a nonmonotone leaf, but yields only slightly better performance than standard algorithms. Moreover, Hu et al. designed a monotonic decision tree algorithm based on rank mutual information (RMI) in 2012 [25] called REMT. Inspired by this idea in [26], Marsala and Petturiti [35] defined rank Gini impurity (RGI) and proposed a binary decision tree classifier (RGMT). Some recent ensemble algorithms have been adapted for monotonic classification [18,23]. These methods have solved some important problems in monotonic classification tasks. In 2015, Qian et al. proposed the fusing monotonic decision tree [38]. Moreover, they discussed attribute reduction and fusion principles. However, most algorithms assume that all features are monotonic with the decision.

Certain features that have a monotonic relationship with the decision are naturally ordered [12,28]; however, some features may be qualitative, such that their relationship with the decision is nonmonotonic. These features are effective in improving classifier performance. In general, features that are monotonically related to decisions are called criteria, whereas other features are referred to as attributes. We provide the example of evaluating whether a credit card is likely to be overdue in Table 1. There are four features and a decision here. Housing loans and overdue history are qualitative variables; the values of these attributes and the decision are not linearly ordered. These features affect overdue ranking. A cardholder with no housing loan is generally likely to be overdue, as is one with a history of being overdue. A lower credit may lead to higher overdue rank; that is, it monotonically decreases with the decision. This makes it necessary to consider the monotonic direction. In Table 1, we see that usage rates monotonically increase with overdue rank.

To evaluate problems such as these, we design partially monotonic decision trees that can improve the handling of attributes and criteria. In this paper, we propose a rank inconsistency rate (RIR) based on rank mutual information to determine whether features are monotonic. Moreover, RIR can capture monotonic directions. Partially monotonic decision trees (PMDT) generate the best split using mutual information (MI) and rank mutual information (RMI). The algorithm handles not only monotonic features, but also considers nonmonotonic features.

The remainder of this paper is organized as follows: In Section 2, we review some related concepts and formulations. In Section 3, we introduce the discriminant feature monotonicity method. In Section 4, we illustrate the construction of PMDTs. In Section 5, we present some experiments used to evaluate the effectiveness of the proposed algorithms. Finally, our conclusions and directions for future work are given in Section 6.

2. Related work

In this section, we review the concepts related to monotonic classification and measuring feature importance in the context of partially monotonic problems.

2.1. Partial monotonicity constraints

Let $\langle U, A, D\rangle$ be a decision table, where $U = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a set of objects, $A = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is a set of attributes to describe the objects, and D is a finite ordinal set of decisions. The value domain of D is $\{d_1, d_2, \ldots, d_k\}$.

Definition 1. Given a set of objects U, $\forall \mathbf{x} \in U$ and $B \subseteq A$, where $B = \{\mathbf{a}_1, \ldots, \mathbf{a}_{m'}\}$. Let $\mathbf{a}_k(\mathbf{x})$ be the attribute value of sample $\mathbf{x}$ on $\mathbf{a}_k$. The ordinal relations between samples in terms of attribute $\mathbf{a}_k$ or D is denoted by $\leq$. Thus, the partial ordering $\preccurlyeq$ on U is defined as

$$\mathbf{x}_i \preccurlyeq_{\mathbf{B}} \mathbf{x}_j \iff \mathbf{a}_k(\mathbf{x}_i) \leqslant \mathbf{a}_k(\mathbf{x}_j), \ \text{for } k = 1, \ldots, m'. \tag{1}$$