Information Sciences 181 (2011) 4959–4968

Contents lists available at ScienceDirect

Information Sciences

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

A PROMETHEE-based classification method using concordance and discordance relations and its application to bankruptcy prediction

Yi-Chung Hu *, Chiung-Jung Chen

Department of Business Administration, Chung Yuan Christian University, Chung-Li 32023, Taiwan, ROC

ARTICLE INFO

Article history:
Received 22 July 2010
Received in revised form 6 May 2011
Accepted 6 June 2011
Available online 8 July 2011

Keywords:
Multiple criteria analysis
PROMETHEE
Preference relation
Ordinal sorting
Bankruptcy prediction

ABSTRACT

Outranking relation theory has been widely used to study pattern classification. Here we propose a classification method with concepts from the flows used in PROMETHEE methods, which are extensively applied in multi-criteria decision aids. PROMETHEE uses a flow, generated on the basis of a preference index and measured by various preference functions for each criterion, to represent the preference intensity for one pattern over another pattern. However, only criteria that are concordant with the preference contribute to a preference index. In the present study, the opinions from discordant criteria are also taken into account. The proposed method newly defines an overall preference index using both concordance and discordance relations for ordinal sorting problems. The final classification decision for a new pattern depends on its net flow. The criteria weights are determined using a genetic-algorithm-based approach. Empirical results obtained for a real-world problem regarding bankruptcy prediction demonstrate that the proposed method performs well compared to other well-known classification methods.

1. Introduction

Pattern classification refers to a problem that predefines classes and then assigns a pattern to one of those classes [23]. Given α ordered classes $C_1, C_2, \ldots, C_\alpha$, $C_1 \succ C_2 \succ \cdots \succ C_\alpha$ indicates that C_1 compromises the most preferred alternatives whereas C_α compromises the least preferred alternatives. Through the development of Elimination and Choice Translating Reality (ELECTRE) methods, the outranking relation theory (ORT) [30–32] has already been used to study classification problems. ORT-based techniques provide preference information using pairwise comparisons between alternatives. Many approaches e.g., [28,33,38] were proposed on the basis of delimiting classes by different fictitious alternatives or reference profiles. The class label of a new pattern is determined on the basis of its comparison to reference profiles. However, determination of reference profiles in real-world situations is often cumbersome [12].

Besides ELECTRE methods, Preference Ranking Organization METHods for Enrichment Evaluations (PROMETHEE) methods introduced by Brans and co-workers [7–10,36] are the other most extensively used ORT techniques [4,12]. The family of PROMETHEE methods is suitable for ranking and choice problems, especially PROMETHEE I and II [1,17,39]. Given two alternatives **a** and **b**, a preference index $P(\mathbf{a},\mathbf{b})$ in the PROMETHEE methods, where P denotes a strict preference relation, is used to measure the strength of the preference for **a** over **b**. The higher the intensity, the stronger is the preference indicated. $P(\mathbf{a},\mathbf{b})$ also represents the flow from **a** to **b**. This means that $P(\mathbf{a},\mathbf{b})$ is a leaving flow for **a** but it is an entering flow for **b**.

Based on PROMETHEE methodology, Figueira et al. [40] proposed PROMETHEE TRI and PROMETHEE CLUSTER to deal with nominal sorting problems by setting reference profiles. These two methods performed classification tasks by computing the

* Corresponding author. Tel.: +886 3 2655130; fax: +886 3 2655199.
E-mail address: ychu@cycu.edu.tw (Y.-C. Hu).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.06.021

single criterion net flows for each pattern instead of overall net flows. Araz and Ozkarahan [2,3] used limiting profiles to delimit different categories to propose the novel PROMSORT method, which assumes that class labels for training samples are unknown. Nevertheless, instead of setting reference profiles, it is feasible to make comparisons between a new pattern and the training samples with known class labels. Previously, PAIRwise CLASsification (PAIRCLAS) [13] developed on the basis of flows in PROMETHEE II was applied to credit risk assessment consisting of two classes, firms with low credit risk (C_1) and firms with high credit risk (C_2). For a new pattern, the higher the net flow derived by leaving and entering flows, the greater is the possibility of assigning it to C_1. However, PAIRCLAS only accounts for flows from C_1 to a new pattern and those from a new pattern to C_2 are taken into account.

Since only criteria concordant with the preference for **a** over **b** contribute to $P(\mathbf{a},\mathbf{b})$, $P(\mathbf{a},\mathbf{b})$ is a measure of the strength of agreement with the proposition "**a** is strictly better than **b**" (i.e., $\mathbf{a}P\mathbf{b}$) without considering the intensity of the indications against $\mathbf{a}P\mathbf{b}$. P can thus be treated as an overall concordance relation. This prompted us to incorporate a discordance relation into P to balance the pros and cons within the set of criteria. The novel classification method proposed here is to construct an overall preference relation P from concordance and discordance relations for ordinal sorting problems. The final classification decision for a new pattern then depends on its net flow. Several concordance/discordance-based classification methods for nominal sorting problems were proposed [41,44], such as PROAFTN [42], PIP and K-PIP [43] and FBI [28]. In comparison with these concordance/discordance-based classification methods which were developed by using reference profiles to characterize each category, the proposed method is quite a different classification approach.

The remainder of the paper is organized as follows. Section 2 introduces the preference index and flows involved in the PROMETHEE methods. Section 3 describes the proposed methodology in detail. In Section 4, Genetic Algorithms (GAs) [16,22,25] are used to develop a genetic-algorithm-based (GA-based) learning algorithm to automatically determine criteria weights that yield high classification performance. Section 5 applies the proposed method to a sample of 57 publicly traded Taiwanese firms in the information and technology industry that failed financially between 2000 and 2008 to identify a financial pre-warning model. The results demonstrate that the proposed method performs well compared to other well-known classification methods. Section 6 presents the discussion and conclusions.

2. PROMETHEE methods

The proposed approach uses concepts from PROMETHEE methods involving the strict preference relation P [28]. The PROMETHEE family includes PROMETHEE I, II, III, IV, V and VI methods. PROMETHEE I gives a partial ranking of the alternatives, version II provides a complete ranking with the net flows, version III defines the preference and indifference relations using the means and deviations for preference indices, version IV deals with a set of infinite alternatives, version V is a procedure for multiple selection of alternatives under segmentation constraints [5], and version VI provides representations of the human brain [6]. PROMETHEE II [9] is considered for the present study since the concept of complete ranking is used to develop the proposed preference model.

Let n be the number of criteria, T be a set consisting of all alternatives and m denote the number of alternatives in T. Each alternative is a vector evaluated by n attributes such that $\mathbf{x}_i$, $\mathbf{x}_j \in T$ can be represented by $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$ $(1 \leqslant i, j \leqslant m)$, respectively. For a generalized preference function $H(s_k)$ for $\mathbf{x}_i$ and $\mathbf{x}_j$ on criterion k, where $s_k = x_{ik} - x_{jk}$ $(1 \leqslant k \leqslant n)$, $\{P, H(s_k)\}$ is called a generalized criterion. Six different generalized criteria (e.g., quasi criterion, level criterion, Gaussian criterion) are usually taken into account. Fig. 1 depicts the Gaussian preference function $H(s_k)$ which is defined as follows:

$$H(s_k) = 1 - e^{\frac{-s_k^2}{2\sigma_k^2}}, \tag{1}$$

where $\sigma_k > 0$ is a preferential parameter that can be determined by the decision-makers, and $s_k = x_{ik} - x_{jk}$. Usually, it is cumbersome for decision-makers to specify a suitable value for σ_k. As for the quasi criterion, the corresponding preference function $H(s_k)$ is defined as follows:

$$H(s_k) = \begin{cases} 0, & \text{if} \quad -q \leqslant s_k \leqslant q, \\ 1, & \text{if} \quad s_k \leqslant -q \quad \text{or} \quad s_k > q. \end{cases} \tag{2}$$

Whereas the preference function $H(s_k)$ corresponding to the level criterion is defined as follows:

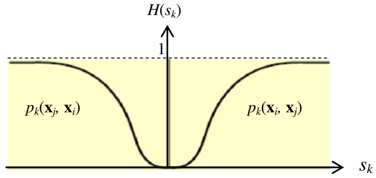

Fig. 1. The Gaussian preference function.