ARTICLE IN PRESS

Technological Forecasting & Social Change xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction

Jamal Ouenniche[a], Blanca Pérez-Gladish[b,*], Kais Bouslah[c]

[a] University of Edinburgh, Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, United Kingdom
[b] University of Oviedo, Faculty of Economics and Business, Oviedo, Asturias 33006, Spain
[c] University of St Andrews, School of Management, Gateway Building, North Haugh, St Andrews KY16 9RJ, United Kingdom

ARTICLE INFO

Keywords:
Out-of-sample prediction
TOPSIS classifier
K-nearest neighbour classifier
Bankruptcy
Risk class prediction

ABSTRACT

Since the publication of the seminal paper by Hwang and Yoon (1981) proposing Technique for Order Performance by the Similarity to Ideal Solution (TOPSIS), a substantial number of papers used this technique in a variety of applications requiring a ranking of alternatives. Very few papers use TOPSIS as a classifier (e.g. Wu and Olson, 2006; Abd-El Fattah et al., 2013) and report a good performance as in-sample classifiers. However, in practice, its use in predicting discrete variables such as risk class belonging is limited by the lack of an out-of-sample evaluation framework. In this paper, we fill this gap by proposing an integrated in-sample and out-of-sample framework for TOPSIS classifiers and test its performance on a UK dataset of bankrupt and non-bankrupt firms listed on the London Stock Exchange (LSE) during 2010–2014. Empirical results show an outstanding predictive performance both in-sample and out-of-sample and thus opens a new avenue for research and applications in risk modelling and analysis using TOPSIS as a non-parametric classifier and makes it a real contender in industry applications in banking and investment. In addition, the proposed framework is robust to a variety of implementation decisions.

1. Introduction

Multi-criteria decision analysis (MCDA) methodologies are widely used for addressing a variety of problems; namely, selection problems, ranking problems, sorting problems, classification problems, clustering problems, and description problems, where selection problems are concerned with identifying the best alternative or a subset of best alternatives; ranking problems are concerned with constructing a rank ordering of alternatives from best to worst; sorting problems are concerned with classifying alternatives into pre-defined and ordered homogenous groups or classes; classification problems are concerned with classifying alternatives into pre-defined and unordered homogenous classes; clustering problems are concerned with classifying alternatives into not pre-defined and not ordered homogenous classes; and description problems are concerned with identifying major distinguishing features of alternatives and perform their description based on these features. In this paper, we are focusing on the solution of classification problems, or equivalently predicting class belonging. To be more specific, we are concerned with the implementation of classifiers and their performance evaluation both in-sample and out-of-sample.

One popular MCDA methodology is Technique for Order Performance by the Similarity to Ideal Solution (TOPSIS) proposed by Hwang and Yoon (1981) and used in many application areas – see Behzadian et al. (2012) for a review including a sample of application areas. This methodology was originally designed for solving ranking problems. In fact, TOPSIS provides a ranking of alternatives based on similarity scores, where the similarity score of each alternative is a function of the distances between the alternative and a couple of benchmarks commonly referred to as the positive and the negative ideal solutions. Later on, TOPSIS has been adapted for solving classification problems. However, to the best of our knowledge, TOPSIS classifiers and their performance evaluation has so far been restricted to in-sample analyses only (e.g., Tansel and Yurdakul, 2010; Abd-El Fattah et al., 2013). In sum, an out-of-sample framework for TOPSIS as a classifier is lacking. The aim of this paper is to fill this gap by proposing a new integrated framework for implementing a full classification analysis; namely, in-sample classification and out-of-sample classification. The proposed framework is intended to make TOPSIS classifiers real contenders in practice and to increase confidence in their use in a variety of critical application areas such as the prediction of risk class belonging (e.g., bankruptcy prediction, distress prediction, fraud detection, credit scoring).

* Corresponding author.
E-mail addresses: Jamal.Ouenniche@ed.ac.uk (J. Ouenniche), bperez@uniovi.es (B. Pérez-Gladish), kbhb@st-andrews.ac.uk (K. Bouslah).

http://dx.doi.org/10.1016/j.techfore.2017.05.034
Received 29 January 2017; Received in revised form 23 April 2017; Accepted 28 May 2017

The remainder of this paper unfolds as follows. In Section 2, we provide a detailed description of the proposed integrated in-sample and out-of-sample framework for TOPSIS classifiers and discuss implementation decisions. In Section 3, we empirically test the performance of the proposed framework in bankruptcy prediction of companies listed on the London Stock Exchange (LSE) and report on our findings. Finally, Section 4 concludes the paper.

2. An integrated in-sample – Out-of-sample framework for TOPSIS classifiers

In the forecasting literature, nowadays prediction models – whether designed for predicting a continuous variable (e.g., the level or volatility of the price of a strategic commodity such as crude oil) or a discrete one (e.g., risk class belonging of companies listed on a stock exchange) – have to be implemented both in-sample and out-of-sample to assess their ability to reproduce or forecast the response variable in the training sample and to forecast the response variable in the test sample, respectively. The rationale behind the necessary implementation and performance evaluation of prediction models both in-sample and out-of-sample lies in the fact that if you feed a properly designed prediction model with some information, it should be able to reproduce/predict that information; therefore, in real life settings, in-sample performance is not enough to quality a prediction model as a good one. Because the future is unknown, out-of-sample implementations and evaluations are used to simulate the future. Out-of-sample implementation and evaluation frameworks are available for parametric prediction models (e.g. statistical models); however, this is not the case for all non-parametric ones (e.g., TOPSIS classifiers).

Hereafter, we shall present our integrated implementation and evaluation framework for TOPSIS classifiers – see Fig. 1 for a graphical depiction of the process. For illustration purposes, we shall customize the presentation of the proposed framework to a bankruptcy application where we reproduce a classical bankruptcy prediction model; namely, the multivariate discriminant analysis (MDA) model of Taffler (1984), within a TOPSIS classifier framework. Recall that Taffler's MDA model focuses on liquidity and makes use of four drivers; namely, Current Assets to Total Liabilities; Current Liabilities to Total Assets; Number of Credit Intervals; and Profit Before Tax to Current Liabilities. Note that lower values are better than higher ones for Current Liabilities to Total Assets and Number of Credit Intervals, whereas higher values of Current Assets to Total Liabilities and Profit Before Tax to Current Liabilities are better than lower ones.

2.1. Input

A training sample $X_E = \{x_{i,j}{}^E; i = 1, \ldots, \#X_E, j = 1, \ldots, m\}$ of cardinality $\#X_E$ and a test sample $X_T = \{x_{i,j}{}^T; i = 1, \ldots, \#X_T, j = 1, \ldots, m\}$ of cardinality $\#X_T$, where each observation i in X_E or X_T is an alternative (e.g., LSE listed firm-year observation) along with a set of relevant features (e.g., bankruptcy drivers) for the analysis under consideration (e.g., Current Assets to Total Liabilities; Current Liabilities to Total Assets; Number of Credit Intervals; Profit Before Tax to Current Liabilities) of cardinality m, and the observed risk or bankruptcy status Y;

2.2. Phase 1: In-sample analysis

Step 1: Choose a *normalization method* (see Table 1) along with a *weighting scheme* w (see Table 2) and use them to transform both training sample data $(x_{i,j}{}^E; i = 1, \ldots, \#X_E, j = 1, \ldots, m)$ and test sample data $(x_{i,j}{}^T; i = 1, \ldots, \#X_T, j = 1, \ldots, m)$ into their normalized counterparts $(r_{i,j}{}^E; i = 1, \ldots, \#X_E, j = 1, \ldots, m)$ and $(r_{i,j}{}^T; i = 1, \ldots, \#X_T, j = 1, \ldots, m)$, respectively, where $x_{i,j}{}^E$ (respectively $x_{i,j}{}^T$) denote the value of feature or driver j of alternative i in the training (respectively, test) sample and $r_{i,j}{}^E$ (respectively $r_{i,j}{}^T$) denote the standardized value of feature j of alternative i in the training (respectively, test) sample.

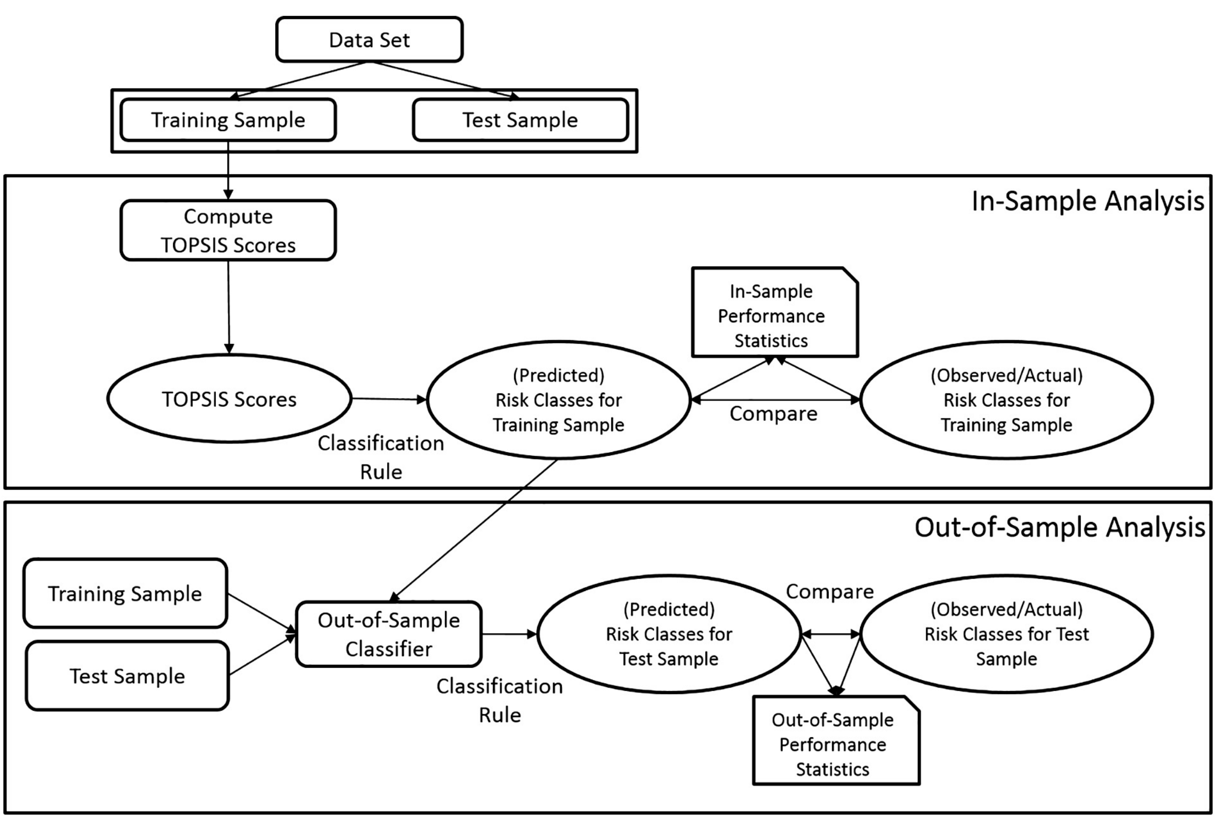

Fig. 1. Generic design of in-sample and out-of-sample analyses of TOPSIS classifiers.