ELSEVIER

Expert Systems with Applications 29 (2005) 879–888

Expert Systems with Applications

www.elsevier.com/locate/eswa

Expositing stages of VPRS analysis in an expert system: Application with bank credit ratings

Benjamin Griffiths, Malcolm J. Beynon*

Cardiff Business School, Cardiff University, Colum Drive, Cardiff, CF10 3EU, Wales, UK

Abstract

The variable precision rough sets model (VPRS) along with many derivatives of rough set theory (RST) necessitates a number of stages towards the final classification of objects. These include, (i) the identification of subsets of condition attributes (β-reducts in VPRS) which have the same quality of classification as the whole set, (ii) the construction of sets of decision rules associated with the reducts and (iii) the classification of the individual objects by the decision rules. The expert system exposited here offers a decision maker (DM) the opportunity to fully view each of these stages, subsequently empowering an analyst to make choices during the analysis. Its particular innovation is the ability to visually present available β-reducts, from which the DM can make their selection, a consequence of their own reasons or expectations of the analysis undertaken. The practical analysis considered here is applied on a real world application, the credit ratings of large banks and investment companies in Europe and North America. The snapshots of the expert system presented illustrate the variation in results from the 'asymmetric' consequences of the choice of β-reducts considered.

Keywords: Bank ratings; Data visualisation; Decision rules; Expert system; VPRS

1. Introduction

Since the introduction of Rough Set Theory (RST) about twenty years ago (Pawlak, 1982, 1991), it has become a popular technique for the classification of objects (Tsumato, Slowinski, Komorowski, & Grzymala-Busse, 2004). Its popularity is a direct consequence of its operational processes, which adhere most closely to the notions of knowledge discovery and data mining (Li & Wang, 2004). These include; operating on the data to identify facts, only from that data which has been utilised and there are no externally imposed assumptions on the data (Jensen, 2004). There is for example no need for normally distributed attribute values as in multivariate discriminant analysis (see Lin & Piesse, 2004). These issues mitigate external concerns placed on a decision maker (DM), moreover they leave the decision maker to undertake their particular analysis (based on a known research theme).

* Corresponding author. Tel: +44 029 2087 5747; fax: +44 029 2087 4419.

E-mail address: beynonmj@cardiff.ac.uk (M.J. Beynon).

0957-4174/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2005.06.008

An illustration of the popularity of RST has been its nascent development (Alpigini, Peters, Skowron, & Zhong, 2002; Tsumato et al., 2004), which has included advances in the areas such as medical applications, bioinfornmatics, image recognition and information retrieval. Here the variable precision rough set (VPRS) approach is considered, which as its name may suggest allows for a level of miss-classification to exist in the decision rules constructed (see Ziarko, 1993a, b; Beynon, 2001). Moreover, central to this study is the exposition of an expert system that undertakes the various stages of VPRS analysis, with emphasis on the appropriate interaction with the DM throughout.

The particular application of the described VPRS expert system, is in the area of bank ratings. Moreover, North American and European banks that have been assigned Moody's Bank Financial Strength Rating (BFSR) are considered (Moody's Europe, 2004). As with the general rating problem, there is a dearth of specific 'public' knowledge on how the credit agencies like Moody's and Standard and Poor's (S&Ps) make their classification decisions (Singleton & Surkan, 1991). This itself has encouraged analysis using a variety of techniques, including multiple regression models (Horrigan, 1966; Molinero, Gomez, & Cinca, 1996; West, 1970;), probit and logit models (Bouzouita & Young, 1998) and neural networks

(Singleton & Surkan, 1991). With the use of VPRS producing sets of readable decision rules then a novel (initial) analysis in this area is also a part of this study.

As an expert system, here its purpose is many-fold; firstly to adequately undertake the necessary analysis (in this case VPRS), secondly to present the results to a DM in a way that benefits them, thirdly control over the analysis is placed fully with the DM. This control issue, with respect to VPRS in particular includes the choice of β-reduct from those identified. Moreover, VPRS is one of a number of techniques that centers around knowledge reduction, the expert system places a level of choice on this reduction with the DM themselves. In the bank rating problem there needs to be details available on the individual banks and their correct or in-correct classification. The snapshots of the VPRS expert system presented in this study elucidate these issues.

The structure of the rest of the paper is as follows: In Section 2, the rudiments of VPRS are described with initial snapshots of the expert system utilized on a small example data set. In Section 3, the bank data set is described. In Sections 4 and 5, the VPRS expert system is applied to the aforementioned bank data set, including snapshots describing the three stages of VPRS outlined previously. In Section 6, conclusions are given as well as directions for future research.

2. VPRS and concomitant expert system (using a small example)

This section briefly describes fundamental aspects of VPRS through a small example and use of the concomitant expert system, for further reading on VPRS and related methodologies see Ziarko (1993a, b), Beynon (2001); Tsumato et al. (2004). The *information system,* central to VPRS, is made up of objects described by sets of condition (**C**) and decision attributes and (**D**). Objects characterized and categorized by the same *condition* and *decision* attribute values are grouped into *condition* (**E**(**C**)) and *decision* (**E**(**D**)) *classes*, respectively.[1] The VPRS expert system described in this study includes the ability to view such an information system, see Fig. 1.

In Fig. 1, the presented information system is made up of seven objects, described by six condition and one decision attribute. The 0 and 1 *descriptor* values[2] enable the construction of condition and decision classes. In summary, the *condition classes* of objects (using all the condition attributes, **C**) as groupings ($\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_4$ and $\mathbf{X}_5$) of indiscernible objects are:

$$\mathbf{X}_1 = \{o_0\},\quad \mathbf{X}_2 = \{o_1, o_4, o_6\},\quad \mathbf{X}_3 = \{o_2\},$$
$$\mathbf{X}_4 = \{o_3\} \text{ and } \mathbf{X}_5 = \{o_5\}.$$

Similarly, the decision classes ($\mathbf{Y}_0$ and $\mathbf{Y}_1$) associated with the decision attribute are:

$$\mathbf{Y}_0 = \{o_3, o_4, o_5, o_6\} \text{ and } \mathbf{Y}_1 = \{o_0, o_1, o_2\}.$$

The relationship (memberships of objects) between these condition and decision classes is fundamental to VPRS (and other RST based methodologies). Subject to a majority inclusion relation with threshold value $\beta \in (0.5, 1]$ (see An, Shan, Chan, Cercone, & Ziarko, 1996), VPRS looks at the condition classes that can be considered associated (classified) with a particular decision class. Moreover, those condition classes which have the property that the *largest group proportion* of objects classified to a decision class is at least β. The proportion of the objects in these classified condition classes is defined as the *quality of the classification*, more formally defined by Ziarko (1993a, b):

$$\gamma^{\beta}(\mathbf{P}, \mathbf{D}) = \frac{\mathrm{card} \cup_{\Pr(\mathbf{Z}|\mathbf{X}_i) \geq \beta} \{\mathbf{X}_i \in \mathbf{E}(\mathbf{P})\}}{\mathrm{card}(\mathbf{U})},$$

where $\mathbf{Z} \in \mathbf{E}(\mathbf{D})$ and $\mathbf{P} \subseteq \mathbf{C}$,

for a specified value of β.[3] An integral part of VPRS is rule construction through the use of *β-reducts* which are particular subsets of condition attributes (**P**) providing the classification of objects with the same $\gamma^{\beta}(\mathbf{P}, \mathbf{D})$ as the full set of attributes **C** (*ibid.*). Formally in VPRS, a β-reduct (RED^{β} (**P**, **D**)), has the twin properties:

(1) $\gamma^{\beta}(\mathbf{C}, \mathbf{D}) = \gamma^{\beta}(RED^{\beta}(\mathbf{C}, \mathbf{D}), \mathbf{D})$,
(2) No proper subset of RED^{β} (**C**, **D**), subject to the same β value can also give the same quality of classification.

Based on an identified β-reduct, a number of different sets of decision rules can be identified (maximal, minimal - see An et al., 1996). Here the minimal rule set is constructed, found through the identification of prime implicants and then further reduction in conditions (descriptor values) in the individual rules (*ibid.*). The details of VPRS presented so far, are its near original form (from Ziarko, 1993a, b), recent developments include those reported in Beynon (2001); Mi, Wu, and Zhang (2004) and Li and Wang (2004). The VPRS expert system exposited here incorporates the work in Beynon (2001), where information over the whole β domain (0.5, 1.0] is utilised. With a possible infinite number of different β values identified to choose from, more than one β-reduct (finite number) may be identified. For the information

[1] These are equivalence classes E(·), where a single member is made up of a group of objects indiscernible with each other based on the attribute values it is associated with (see Beynon, 2001).

[2] VPRS works on a level of discretised data (rather than continuous values), hence may require the lowering of the level of granularity of the considered data set (see Beynon, 2004).

[3] The full description of VPRS involves the identification of certain approximation regions (positive, boundary and negative), each containing groups of objects, see Beynon (2001) and Ziarko (1993a, b). For brevity this paper purposely limits the level of formal notation described herein.