Applied Soft Computing 71 (2018) 835–848

Contents lists available at ScienceDirect

Applied Soft Computing

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

# A cautious ranking methodology with its application for stock screening

Peng Song [a,b], Jiye Liang [b,c,*], Yuhua Qian [b,c], Wei Wei [b,c], Feng Wang [b,c]

[a] School of Economics and Management, Shanxi University, Taiyuan 030006, Shanxi, China
[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, Shanxi, China
[c] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China

ARTICLE INFO

*Article history:*
Received 12 March 2017
Received in revised form 29 April 2018
Accepted 3 July 2018

*Keywords:*
Interval data
Ranking decision
Rough sets
Information entropy
Stock screening

ABSTRACT

To reasonably and effectively solve the ranking decision problem, there currently exist various meaningful methods that are focused on different perspectives. However, the ranking decision problem is a systematic issue that involves data representation, the dominance relation, feature selection, and ranking mechanism. In this study, we aimed to build a novel ranking methodology by taking into account both the inherent multicriteria nature of practical decision situations and cautious decision makers' preferences. In order to better reveal the entirety of the data set, the form of interval data is introduced to characterize the ranges of attribute values. For the purpose of improving the decision performance, we develop a measurement called interval ordered conditional entropy to extract the most representative condition attributes having significant ordered relevance to the decision attribute. Based on the cautious dominance relation introduced for interval data, a two-step ranking mechanism with cautious characteristics is introduced that utilizes an interval ordered information table organized according to the previously selected informative attributes. In addition, the validity of this ranking method is tested through a detailed case study on stock screening decisions involving three successive rounds of tests. The corresponding results indicate the effectiveness of the methodological approach proposed in this paper.


## 1. Introduction

In practical decision environments, ranking decision problems, such as that involved in ranking universities [1,2], venture investment [3], and market segment evaluation [4], is a typical type of decision problem. Thus far, many effective and reasonable ranking methods have been constructed, including TOPSIS [5,6], ELECTRE [7], the analytic hierarchy process (AHP) [6,8], etc. It should be noted that, in the past decades, the forms of data representation appear to have become more complex because of the increasingly intricate and uncertain decision situations. Accordingly, many researchers have paid considerable attention to ranking approaches in the context of interval-valued data [3,9–11].

In realistic decision problems, it is increasingly found that the decision behaviors of decision makers act as important influence factors in their decision making. Yao built the theory of three-way decisions, which effectively characterized the common human behavior of trisecting a universal set in problem solving [12,13]. Wu and Chiclana proposed new attitudinal expected score and accuracy functions for ranking interval-valued intuitionistic fuzzy numbers that take into account the characteristics of decision makers' risk attitude [14]. Ruan and Shi integrated interval comparison techniques into scenario analysis methods for monitoring and assessing fruit freshness in an IOT-based e-commerce delivery system [15]. A review of the existing literature shows that the issue of ranking interval data considering decision behaviors has become a promising and interesting research field [14–18]. In reality, risk aversion is a typical type of human decision behavior [19,20]. In particular on the background of exacerbated market risk after the global financial crisis in 2008 [21], the study of risk aversion-based ranking models is of significance and value. Chen and Zou developed a new group decision-making method to solve the supplier selection problem, in which the weights of the decision makers are appropriate for risk avoiders under an uncertain environment [22]. Through introducing the hyperbolic absolute risk aversion utility function, Gao et al. built a new operator for group decision making that can reflect the decision makers' risk attitude [23]. The above methods

* Corresponding author at: Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, Shanxi, China.

*E-mail addresses:* songpeng@sxu.edu.cn (P. Song), ljy@sxu.edu.cn (J. Liang), jinchengqyh@sxu.edu.cn (Y. Qian), weiwei@sxu.edu.cn (W. Wei), sxuwangfeng@126.com (F. Wang).

https://doi.org/10.1016/j.asoc.2018.07.004
1568-4946/

effectively reflect the decision behavior of risk averse decision makers from their specific perspectives. However, the cautious ranking decision problem is a systematic issue. Therefore, a methodological approach that considers risk aversion behavior in each step of the problem solving (e.g., in the dominance relation, attribute weights, and ranking mechanism) is desirable in the context of interval data.

In general, special ranking decision issues have specific decision targets. In fact, in certain important and classical ranking decision problems, such as stock screening and venture investment, feature selection is a key preprocessing step of the ranking methodology [3,24]. Given the nature of the aforementioned decision problems, the basic issue is to discover the informative indicators that have significant ordered relevance to the stock return or investment return. Through using an effective and reasonable feature selection algorithm, we can obtain the pertinent evaluation attributes, and then effectively provide a ranking result for improving the investment performance [25]. Obviously, feature selection is an indispensable component in the overall ranking procedure in the type of decision situation mentioned above. It should be noted that attribute reduction in rough set theory, introduced by Pawlak [26,27], has become a useful tool that is widely used for feature selection [3,28–34]. Ziarko defined the notion of $\beta$-reduct and provided a battery of attribute reduction approaches in the variable precision rough set model [30]. To address the problem of time complexity in dynamic datasets, Chen et al. proposed an incremental algorithm for attribute reduction in the variable precision rough set model [31].

The research outcomes mentioned above provided a wide variety of effective feature selection algorithms for problem solving in different complex and uncertain situations. It should be emphasized that, to handle ranking decision problems with interval data effectively, Qian [3] presented an attribute reduction approach based on a discernibility matrix. In fact, discernibility matrix-based approaches constitute a typical type of attribute reduction method in rough set theory [35]. However, they have a disadvantage in that the discernibility function does not fully correspond with the specific decision target. Basically, given the nature of ranking decisions, feature selection should consider the ordered relevance between the key evaluation feature subsets and the decision target. Consequently, the exploration of the feature selection approach that takes into account the characteristic of ordered relevance has naturally become a relevant research issue.

In the area of practical decision situations, the issue of stock screening can be treated as a ranking decision problem [36,37] and many valuable multiple criteria decision making (MCDM) methods are also constantly emerging [7,38,39]. Xidonas et al. presented an ELECTRE Tri method, in which three types of investors' preferences, consisting of the conservative, balanced, and aggressive investment profiles, are considered in the decision-making process of stock selection [38]. Considering both the preferences involved in evaluating the relevant criteria and the decision behaviors of investors, Sevastjanov and Dymova proposed a cautious stock ranking strategy that considers the high level coincidence between a firm's financial performance and its market success in the context of fuzzy-valued data [37]. Indeed, the strategy is quite credible, because the proposed program can reject "unsafe" firms, the market success of which may be derived from the subjective experts' opinions, rumors, or other factors. However, two important problems related to this cautious strategy remain to be considered, one of which is the determination of an appropriate membership function of fuzzy subsets, which remains a challenging issue that needs to be addressed [36,37]. Indeed, in practical issues, interval data have become a significant type of data representation form for better revealing the entirety of a data set [40–42]. In particular for stock screening decisions, most indicators always show the characteristic of fluctuations. In fact, the form of interval data has a better capability to reflect this type of uncertainty [40,41]. Therefore, in this study it was considered desirable that the representation of interval data be introduced into the decision process. The second problem is the determination of the informative financial indicators that have significant ordered relevance to stock return in order to improve investment performance by means of using an effective and reasonable feature selection algorithm [25,43]. It is considered that feature selection is the key process in the stock screening decision [44]. Overall, the stock screening decision is a systematic issue that involves a series of successive problems, including data representation, feature selection, and ranking mechanism.

In summary, the objective of this study was to develop a novel ranking methodology that involves interval data representation, feature selection, and a ranking mechanism in the context of a risk averse situation. This was the motivation of our research. In this study, our objective was to solve three key problems:

- How can a systematic cautious ranking method for interval data be constructed?
- How can an ordered-relevance-preserving feature selection method in the proposed cautious ranking methodology be built?
- How can stock screening decision problems be solved by employing the proposed cautious ranking methodological approach?

To build a novel cautious ranking methodological approach, in this study, we first introduced a cautious dominance relation into the initial decision table as the fundamental representation for depicting ordered information. Then, on the basis of interval data transformation for all the attributes and a subsequent sorting process for the decision attributes, we were able develop a new interval ordered decision table (IODT). Based on the previous steps, a measurement of ordered relevance, called interval ordered conditional entropy, for extracting the representative feature subsets was devised. Finally, in terms of the above selected pertinent features, we introduce a cautious ranking mechanism to obtain the ranking results. Simultaneously, in the process of decision modeling analysis, the validity of the proposed approach was verified by a case study on stock screening. The proposed methodology can further improve the investment performance of stock screening decisions. The flow chart of cautious ranking methodology is shown in Fig. 1.

The remainder of this paper is organized as follows. In Section 2, we review some preliminary notions and definitions related to the dominance relation, information entropy in rough set theory, and ranking mechanisms. In Section 3, we describe the building of a systematic cautious ranking methodology that includes a sequence of decision processes, namely, a cautious dominance relation for interval data, interval data-based data representation, a feature selection method, and a two-step ranking mechanism. In Section 4, we present a detailed case study of stock screening decisions to verify the effectiveness of the methodology proposed in this paper. Finally, Section 5 concludes this paper by presenting some remarks and discussions.

## 2. Preliminaries

An *information system* (IS) is a quadruple $S=(U, AT, V, f)$, where $U$ denotes a finite non-empty set of objects and $AT$ an attribute set, $V=\bigcup_{a\in AT}V_a$ and $V_a$ represents a domain of attribute $a$, and $f: U\times AT\rightarrow V$ is a total function such that $f(x, a)\in V_a(a\in AT, x\in U)$, namely, an information function [27,45]. An IS is called an *interval information system* (IIS) if $V_a$ is a set of interval-valued numbers, where $f(x, a)$ is denoted by

$$f(x,a)=[a^L(x),a^U(x)]=\{p \mid a^L(x)\leqslant p\leqslant a^U(x), a^L(x), a^U(x)\in \mathbf{R}\}.$$