Information Sciences 595 (2022) 89–118

Contents lists available at ScienceDirect

Information Sciences

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

A regret theory-based three-way decision approach with three strategies

Jinxing Zhu, Xueling Ma *, Jianming Zhan

School of Mathematics and Statistics, Hubei Minzu University, Enshi, Hubei 445000, China

ARTICLE INFO

Article history:
Received 17 October 2021
Received in revised form 13 February 2022
Accepted 16 February 2022
Available online 22 February 2022

Keywords:
Three-way decision
Dominance relation
Regret theory
Multi-attribute decision-making

ABSTRACT

Breast cancer is a malignant tumor that seriously threatens women's health. Although classic multi-attribute decision-making (MADM) techniques can handle this kind of medical problem, a decision-maker (DM) can only collect sample data under various indicators for result ranking and analysis. The three-way decision (3WD) theory further supplements a classification scheme. In addition, a DM's own limited rationality and personality traits have a strong impact on decision-making results. By using a dominance-based rough set approach (DRSA), a new 3WD method based on the regret theory (RT) with optimistic, neutral and pessimistic strategies (3WD-RT-OEP) is constructed in a fuzzy environment, so as to prevent diseases in advance and improve the survival rate of patients. First, attribute weights are calculated based on the similarity between the defined attributes of dominance classes, and a new method for calculating conditional probabilities is proposed to enhance the objectivity of the method by the similarity between object dominance classes and decision classes. Second, the score functions of three strategies are proposed by considering regret and rejoicing values in RT, and specific steps and algorithms of the 3WD-RT-OEP method are given as well. Finally, the validity and rationality of the constructed method are proved by experimental analysis with the support of case studies and supplementary data sets.

1. Introduction

As one of the most common malignant tumors, breast cancer seriously affects women's physical and mental health and even endangers their lives. In 2020, 2.3 million women were diagnosed with breast cancer and 685,000 cases were died. By the end of 2020, 7.8 million living women had been diagnosed with breast cancer in the past five years, making it the most common cancer in the world. In order to reduce the number of disability-adjusted life years lost to breast cancer in women, Song et al. [29] used impulse-radio-ultra-wideband-radar detectors to detect breast tissue after resections. Feng et al. [3] used only the node size of three treatments to predict the likelihood that chemotherapy would respond to breast cancer. Clinical practice over the years has proved that cancer is not an incurable disease, but the key is early detections and treatments. Although there have been various improvements for treatment methods of breast cancer in recent decades, the death rate has not been significantly reduced, and the main reason for this is the late diagnosis. This requires that we advocate to detect cancer early to reduce the emergence of advanced cancers. With the rapid growth of Internet data, the application of data mining and other technologies in medical treatments can effectively improve the productivity and medical efficiency of

* Corresponding author.
E-mail addresses: venus0530@163.com (J. Zhu), zjmhbmy@126.com (X. Ma), zhanjianming@hotmail.com (J. Zhan).

https://doi.org/10.1016/j.ins.2022.02.030
0020-0255/

the medical industry. Traditionally, breast cancer is identified by specialists, and inconspicuous early features may delay a patient's condition. Thus, the digital analysis of patients' relevant detection indicators is an effective way to diagnose the benign and malignant nature of breast cancer and improve breast cancer survival rates.

1.1. A brief review of MADM

The MADM method [20] integrates and ranks multiple options via considering multiple criteria, i.e., physical measurement indexes, so as to select the best scheme, which is just suitable for solving such problems by data detections of multiple indicators in breast cancer samples and then ranking patients according to the degree of diseases. The application of MADM methods in medical diagnosis has been presented in diverse fields [8,2,49,20]. Although MADM methods can rank schemes by integrating criteria values, they ignore the possible losses or benefits of different actions in decision-making processes. In addition, the majority of MADM methods are two-way decision (2WD), i.e., there are only two decision-making results: acceptances and rejections. When the information on the object is comprehensive and obvious, the method can directly give the object an acceptance or rejection decision outcome, however this is difficult to accurately implement in increasingly complex decision environments.

In addition, machine learning algorithms have been rapidly developed during the past decades, which can be roughly classified into two branches, i.e., supervised learning algorithms and unsupervised learning algorithms [31]. In specific, supervised learning algorithms mainly include linear regression, logistic regression, neural networks, support vector machine, random forest, etc. These models demand sufficient and compatible training samples to ensure appropriate training of models. The current paper primarily focuses on MADM problems, and existing related data sets do not include sufficient and compatible training samples, thus supervised learning algorithms can not efficiently address MADM problems. For unsupervised learning algorithms, the representatives are clustering, principle component analysis, matrix factorization, etc. Due to the structure of MADM information systems, clustering can be used to seek the inherent distribution structure of data and try to divide the samples in the data set into several disjoint subsets. Commonly used algorithms such as centroid algorithms [6] are sensitive to noise and isolated point data, hierarchical clustering method may lead to clustering results that differ from the true class distribution due to a limit on the number of clustering features per node, and different combinations of parameters in density clustering [6] also have a large impact on final clustering results. For MADM problems, a DM needs to choose from a large number of alternatives and determine the ultimate goal, only classifying alternatives cannot meet a DM's needs. Meanwhile, uncertain multi-classification problems are not applicable to some practical problems. For instance, when a doctor is diagnosing whether a patient needs to receive treatment, multi-classifying the patient only increases the decision burden and reduces the efficiency of a doctor's diagnosis. In light of the above perspectives, due to the strong explainable ability from the aspect of cognitive sciences, it is widely recognized that 3WD can effectively address MADM problems.

1.2. A brief review of 3WD

3WD is created as a bridge to the application of the rough set theory to practical decision-making problems. As a mathematical tool to portray incomplete, uncertain knowledge and data generalization, rough sets can effectively analyze and deal with uncertain information such as imprecision, inconsistency and incompleteness. Since 1982, Pawlak [24] pioneered the basic concept of rough sets for the idea of boundary line regions, many scholars have extended the Pawlak's rough set model from different perspectives [23],[10]. Hu et al. [5] introduced a feature selection method based on mutual information. Further, since classic rough sets can only deal with information systems with discrete attributes based on equivalence relations, they will cause some losses to decision information tables with biased order relations. In view of this, Greco et al. [4] put forward the theory of DRSA instead of an indistinguishable relation, which has been widely used in customer relationship managements [16], fault detections [28], etc.

However, both the classic rough set theory and the DRSA, which replace an equivalence relation with a dominance relation, do not consider the fault tolerance of uncertain objects. In order to perfect the classic rough set theory, Yao [45] constructed decision-theoretic rough sets (DTRSs) in combination with the Bayesian decision process and established 3WD on this basis. 3WD is also widely used as a decision-making method in practical problems. As an illustration, in medical diagnosis processes, when a patient's clinical test information is sufficient, a doctor can immediately make a decision to treat a patient with the disease or refuse treatment without the disease; when a patient's clinical test information is not sufficient to fully confirm the diagnosis, a doctor will further observe and diagnose a patient and make a decision when the information is sufficient.3WD is a relatively young subject with strong universality. Recently, in terms of theoretical models, Yao [44] established a wide 3WD idea. The basic geometric concepts of points, lines, triangles and circles, as well as the more complex structures derived from them, are used to physically explain ternary thinking, ternary calculation and ternary processing. Based on the membership function and non-membership function of intuitionistic fuzzy sets, Yang et al. [41] solved a pair of thresholds by loss minimizations, divided the universe into three parts and constructed a three-way approximation model. Qian et al. [26] proposed hierarchical sequential 3WD models by combining sequential 3WD with hierarchical rough set models. In terms of method researches, Liu et al. [19] proposed a general optimization model based on the Karush–Kuhn–Tucker condition to determine the threshold pairs in DTRSs with different semantics. Yang et al. [40] explored the method of constructing shadowed sets from the perspective of 3WD. Liu et al. [18] updated 3WD with group consensus by using convex