Knowledge-Based Systems 39 (2013) 159–172

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Structure learning for belief rule base expert system: A comparative study

Leilei Chang [a], Yu Zhou [a], Jiang Jiang [a], Mengjun Li [a,*], Xiaohang Zhang [b]

[a] Department of Management, National University of Defense Technology, Changsha 410073, Hunan, PR China
[b] Department of System & Software Engineering, 63891, Luoyang 471000, Henan, PR China

ARTICLE INFO

Article history:
Received 25 July 2012
Received in revised form 21 October 2012
Accepted 23 October 2012
Available online 16 November 2012

Keywords:
Belief rule base
RIMER
Structure learning
Dimensionality reduction
Principle component analysis

ABSTRACT

The Belief Rule Base (BRB) is an expert system which can handle both qualitative and quantitative information. One of the applications of the BRB is the Rule-base Inference Methodology using the Evidential Reasoning approach (RIMER). Using the BRB, RIMER can handle different types of information under uncertainty. However, there is a combinatorial explosion problem when there are too many attributes and/or too many alternatives for each attribute in the BRB. Most current approaches are designed to reduce the number of the alternatives for each attribute, where the rules are derived from physical systems and redundant in numbers. However, these approaches are not applicable when the rules are given by experts and the BRB should not be oversized. A structure learning approach is proposed using Grey Target (GT), Multidimensional Scaling (MDS), Isomap and Principle Component Analysis (PCA) respectively, named as GT–RIMER, MDS–RIMER, Isomap–RIMER and PCA–RIMER. A case is studied to evaluate the overall capability of an Armored System of Systems. The efficiency of the proposed approach is validated by the case study results: the BRB is downsized using any of the four techniques, and PCA–RIMER has shown excellent performance. Furthermore, the robustness of PCA–RIMER is further verified under different conditions with varied number of attributes.

1. Introduction

The Belief Rule Base (BRB) system is developed from the traditional IF–THEN rule base to represent different types of knowledge under uncertainty [1–3], which is indispensable when analyzing problems involving human decisions [4]. One of the most important applications of the BRB is the RIMER approach [5–7], which is short for the Rule-base Inference Methodology using the Evidential Reasoning. Using a BRB, RIMER is able to handle different kinds of information under uncertainty, qualitative and/or quantitative, linguistic and/or numerical, complete and/or incomplete. The applications of RIMER include the Multiple Attribute Decision Analysis (MADA) problem [8], group decision making [9,10], risk analysis [11,12], trade-off analysis [13], pipeline leak detection [14], military capability evaluation [15], etc.

However, when the BRB is applied in RIMER, it is required to cover all the possible combinations of each alternative for each attribute. This is, in fact, a combinatorial explosion problem [16]: the size of the BRB would grow exponentially along with the increase of the number of the attributes and/or the alternatives for each attribute. Therefore, there is a pressing demand to downsize the BRB.

From the perspective of its origination, the BRB is classified as physical-model-based and expert-knowledge-based. For the physical-model-based BRB, the number of the attributes is fixed while the alternatives for each attribute are changeable; for the expert-knowledge-based BRB, only the attributes are changeable.

There are four parameters concerning the structure of a BRB: the number of the attributes, the alternatives for each attribute, the number of the scales and the alternatives of each scale in the assessment result [17]. The first two parameters determine the size of a BRB, which leads to two means to downsize a BRB. To downsize a BRB by means of identifying either parameter is called the structure learning of a BRB. Also, to identify the alternatives of each attribute is called parameter learning of a BRB as well. Since the four parameters are the most important feature of a BRB concerning its structure, to perform structure and/or parameter learning is to perform feature selection on a BRB.

Different techniques are applicable for feature selection in a BRB. Zhou [18,19] used hidden Markov chain to model the correlation between environmental variables and observable variables, based on which a parameter learning approach was proposed; Jiang [16] proposed a parameter model of Evidential Network (EN) in which the Conditional Belief Function (CBF) and the BRB model were combined, furthermore Jiang conducted the parameter learning study within the framework of EN. Tsai [20,21] proposed a rule mining algorithm in the presence of concept drift, and the idea of feature extraction was demonstrated in a patient diagnose case

* Corresponding author. Tel.: +86 0731 84574569.
E-mail address: mjli11260744@gmail.com (M. Li).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.10.016

study context when he managed to accurately discover the underlying governing rules [21]. Rule induction [22,23] mainly focused on evaluating rules (of a BRB) so that the feature of the original rule set could be completely extracted and preserved while the number is considerably reduced. In fact there were over 40 measures in related literatures [24,25]. Suzuki [26] summarized that there was no single universal measure for all rule set analysis. Back to the framework of RIMER, Zhou [18,19] proposed the concept of "statistical utility", which resembled the evaluation measure in the previously introduced rule induction. Zhou used "statistical utility" (1) to rate a rule, (2) to determine whether a rule should be kept by setting a valve value for the "statistical utility", and (3) to learn what the alternatives for each attribute really were.

There has been extensive research on the structure learning for the physical-model-based BRB systems. Yang [27] proposed the first generic BRB learning framework with an optimization model for training the BRB systems. Xu [14] also proposed a training methodology for the BRB systems and applied it in a pipeline leak detection case study. Later more studies were carried out on learning and training the BRB either in a theoretical context and/or a practical context [17,28]. Zhou argued that those optimization models were all offline and locally optimal, which were not suitable for a dynamic fashion [29,30]. Therefore, Zhou [31,32] proposed an online updating BRB approach, which did not require to collect a complete set of data before training the BRB.

However, for the expert-knowledge-based BRB systems, these approaches are no longer applicable. Because the alternatives for each attribute are pre-set and fixed while the number of the attributes is changeable. Therefore, another way is taken to solve this problem: to reduce the number of the attributes. Numerous approaches and/or techniques can help to meet this challenge, of which four techniques are applied in this study: the Grey Target (GT) theory, the Multidimensional Scaling (MDS), the Isomap and the Principal Component Analysis (PCA). MDS, Isomap and PCA are all dimensionality reduction techniques. Along with GT, they are all typical means for feature selection [33–36].

GT, as a non-statistical technique, is part of the Grey System Method (GM) which was first introduced by Deng in early 1980s [33]. Since then GT had become more and more popular in dealing with problems that have partially unknown factors, or "grey" information [34,35]. Successful applications of GT include pattern recognition, feature selection, technical assessment, etc. [37,38]. Compared with traditional statistical approaches, GT requires much less information to estimate the behavior of the unknown systems. The "approaching degree" [33] in GT is used to select key attributes in this study.

MDS [39,40] and Isomap [41] are two nonlinear techniques for dimensionality reduction. MDS [42] uses a "stress function" to measure the geometric distances between the original high-dimensional space and the transformed low-dimensional space, which represents the quality of the mapping. Isomap [41] takes consideration of the distribution of the neighboring datapoints by preserving the pairwise geodesic distance between datapoints, which is the distance between two points measured over the manifold. The two techniques have become more and more poplar so that there have been several computer-based packages [43]. It should be noted that there are many other nonlinear dimensionality reduction techniques, such as LLE [44] and kernel PCA [45].

PCA is the most representative linear dimensionality reduction technique. PCA was commented as a "multivariable technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables" [46]. So far, PCA has become the most widely used and popular linear dimensionality reduction technique [36], which is traditionally used to extract condensed and representative information from massive information [47,48]. The input of PCA can be either statistical data or standardized experts' knowledge. PCA requires certain dependency on the attributes that forms the "data table": the more dependent (less independent) the attributes (as input) are, the better representative the principal components (as output) are.

The dimensionality reduction approaches (including PCA as the representative linear approach) are interrelated and even equivalent in certain cases. For example, the earliest MDS used PCA to analyze correlated data by treating them in the Euclidean distances [49].

The remainder of this study is organized as follows. The RIMER approach and the demand to downsize a BRB are introduced in Section 2. The structure learning approach is proposed in Sections 3 and 4 using GT, MDS, Isomap and PCA respectively. In Section 5, a case to evaluate the overall capability of an Armored System of Systems is studied to validate the efficiency of the proposed approach. This study is concluded in Section 6.

2. Problem demonstration

RIMER [5–7] was proposed by Yang based on the evidence theory [50,51]. RIMER consists of two parts, first building the BRB, and then integrating the activated rules from the BRB using the ER algorithm, as briefly introduced in Sections 2.1 and 2.2. When a BRB is constructed, it is required to cover all the possible combinations of each alternative for each attribute. When there are too many attributes and/or too many alternatives for each attribute, the size of a BRB would grow exponentially, which is a combinatorial explosion problem as demonstrated in Section 2.3.

2.1. Belief rule base

The BRB is applied in RIMER to reflect the system dynamics, such as qualitative and/or quantitative, complete and/or incomplete, linguistic and/or numerical information. A BRB is comprised of a series of belief rules. The kth rule is described as:

$$R_k: \ if \quad A_1^k \wedge A_2^k \wedge \cdots \wedge A_M^k, \quad then\{(D_1,\beta_{1,k}),\ldots,(D_N,\beta_{N,k})\} \tag{1}$$

where $A_i^k (i = 1,\ldots,M;\ k = 1,\ldots,L)$ represents the status of the ith attribute, M represents the number of the attributes, N represents the number of the scales, L represents the number of the rules in the BRB, $\beta_{j,k}$ represents the belief for the jth scale, D_j. The kth rule is complete when $\sum_{j=1}^{N}\beta_{j,k} = 1$, and the kth rule is incomplete when $\sum_{j=1}^{N}\beta_{j,k} < 1$.

2.2. Rules integration using the ER algorithm

2.2.1. Weights for activated rules

The activated weight for the kth rule is

$$\omega_k = \frac{\theta_k \prod_{i=1}^{M} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{M} (\alpha_i^l)^{\bar{\delta}_i}}, \quad \text{and} \quad \bar{\beta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M}\{\delta_i\}} \tag{2}$$

where θ_k represents the comparative weight of the kth rule, δ_i represents the weight of the$i$th attributes in the kth rule, and α_k represents the matching degree of the input with the kth rule. When $\alpha_k = 0, \omega_k = 0$, the kth rule is not activated.

2.2.2. Adjustments for belief distribution

When the input information is incomplete, certain adjustments are made upon the belief distribution of each scale in the conclusion part as following:

$$\bar{\beta}_{lk} = \beta_{lk}\mu_k \tag{3}$$

where