Knowledge-Based Systems 53 (2013) 129–141

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A novel belief rule base representation, generation and its inference methodology

Jun Liu [a,c,*], Luis Martinez [b], Alberto Calzada [a], Hui Wang [a]

[a] School of Computing and Mathematics, University of Ulster, Northern Ireland, UK
[b] Department of Computer Science, University of Jaen, Jaen, Spain
[c] School of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan, China

ARTICLE INFO

Article history:
Received 30 May 2013
Received in revised form 13 August 2013
Accepted 16 August 2013
Available online 30 August 2013

Keywords:
Rule-based system
Belief distribution
Uncertainty
Decision-making
Inference mechanisms
Evidential reasoning approach

ABSTRACT

Advancement and application of rule-based systems have always been a key research area in computer-aided support for human decision making due to the fact that rule base is one of the most common frameworks for expressing various types of human knowledge in an intelligent system. In this paper, a novel rule-based representation scheme with a belief structure is proposed firstly along with its inference methodology. Such a rule base is designed with belief degrees embedded in the consequent terms as well as in the all antecedent terms of each rule, which is shown to be capable of capturing vagueness, incompleteness, uncertainty, and nonlinear causal relationships in an integrated way. The overall representation and inference framework offers a further improvement and great extension of the recently developed belief Rule base Inference Methodology (refer to as RIMER), although they still share a common scheme at the final step of inference, i.e., the evidential reasoning (ER) approach is applied to the rule combination. It is worth noting that this new extended belief rule base representation is a great extension of traditional rule base as well as fuzzy rule base by encompassing the uncertainty description in the rule antecedent and consequent. Subsequently, a simple but efficient and powerful method for automatically generating such extended belief rule base from numerical data is proposed involving neither time-consuming iterative learning procedure nor complicated rule generation mechanisms but keeping the relatively good performance, which thanks to the new features of the extended rule base with belief structures. Then some case studies in oil pipeline leak detection and software defect detection are provided to illustrate the proposed new rule base representation, generation, and inference procedure as well as demonstrate its high performance and efficiency by comparing with some existing approaches.

1. Introduction

Among many varieties for knowledge representation, it is widely recognized that the rule based one is one of the most common frameworks for expressing various types of knowledge in an intelligent system [22,1]. Taking advantage of the beauty of representing and manipulating human knowledge at first, in the design and implementation of advanced rule-based systems for supporting human decision making, it is desirable to endower the rule-based system with certain representation scheme and processing capabilities to handle simultaneously vagueness, incompleteness and uncertainty in conjunction with the flexibility in incorporating different types of input data formats, such as numerical, interval, uncertain value, or even subjective judgments.

During the last quarter of a century many different types of rule-based systems emerged, certainly including the fuzzy rule-based system [33,25], which, as one of the dominant and main frameworks in rule-based systems, has been widely accepted, investigated and applied in many application areas. Moreover, in recognition of the need to handle hybrid information with uncertainty in human decision making, a new methodology has been proposed recently for modeling a hybrid rule base using a belief structure and for inference in the belief rule-based system using the evidential reasoning (ER) approach [29]. The methodology is referred to as a belief Rule base Inference Methodology using the Evidential Reasoning approach – RIMER [30], where a rule base is designed with belief degrees embedded in the consequent term of a rule, called a *belief rule base* (BRB), is used to capture nonlinear causal relationships as well as uncertainty. The inference of the belief rule based system is implemented using the ER approach, this has been a distinct feature compared with the existing rule based inference methodologies. RIMER approach has been further investigated and its results and relevant extensions have proved to be highly

* Corresponding author at: School of Computing and Mathematics, University of Ulster, Northern Ireland, UK. Tel.: +44 2890368919.
E-mail address: j.liu@ulster.ac.uk (J. Liu).

0950-7051/$ - see front matter
http://dx.doi.org/10.1016/j.knosys.2013.08.019

positive in solving decision problems cross different application areas, such as, among others, safety and risk analysis, oil pipe leak detection and some other application in health care and engineering systems [31,32,24,2,17,16,15,14,13,12,19,18,6,21,10,11,34–37].

Among other issues, how to generate the rule base is a fundamental issue when designing and implementing a rule base system. Hence, various methods were proposed for automatically generating the rule base from sample data set. Most of these methods have involved iterative training algorithm or complicated generation scheme, e.g., gradient descent learning methods, genetic-algorithm-based methods, least-squares methods, a fuzzy c-means method, and a fuzzy-neuro method for learning fuzzy rule base, however they are either time consuming or using complicate rule generation strategies with the need of additional learning tool. For the RIMER approach, an optimal modeling has been proposed [31,17], with some further improved learning algorithms [34–36].

In this paper, to facilitate the more general and advanced application cases to handle simultaneously imprecision, incompleteness and uncertainty in conjunction with the flexibility in incorporating different types of input data format, as well as more flexible and simpler rule base generation scheme, the belief rule base in RIMER is extended with belief degrees embedded also in the entire antecedent terms of each rule. Most importantly, a simple but efficient and powerful method for automatically generating such extended belief rule base from numerical data is proposed, which is mainly attributed to the new features of the extended belief rule base. The main advantages of which over most of traditional learning approaches is its simplicity and efficiency because it involves neither time-consuming iterative learning procedure nor complicated rule generation mechanisms, but keep the relatively good performance. The work has been briefly outlined in [13]. This paper aims at extending, refining, completing and systematizing the results in [13].

The rest of this paper is organized as follows. Extended belief rule base and its inference framework are proposed in Section 2, including a brief overview of belief rule base in the RIMER approach. In Section 3, we propose a simple but efficient extended belief rule generation method with no time-consuming iterative procedure. Section 4 discusses and proves extended belief rule base inference system as a universal approximator, which shows the soundness of the methodology and provides the theoretical basis for successful applications of this new methodology to many different practical problems. The proposed rule representation, generation, and inference scheme as well as its performance are demonstrated in Section 5 using some case studies in oil pipe leak detection as well as in software defect prediction compared with some existing approaches. Conclusions are drawn in Section 6.

2. Extended belief rule based inference methodology

This section introduces the extended belief rule base along with its inference procedures.

2.1. Extended belief rule base

The belief rule base (BRB) introduced in RIMER approach [30] is summarized firstly in this section, which is designed with belief degrees embedded in the entire consequent terms. Then it is extended into a new belief rule base with belief degrees embedded in the consequent terms as well as in the all antecedent terms of each rule.

Suppose a BRB is given by $\boldsymbol{R} = \{R_1, R_2, \ldots, R_L\}$ with the kth rule represented as follows [30]:

$$R_k:\ \text{IF } \boldsymbol{U} \text{ is } \boldsymbol{A}^k \text{ THEN } V \text{ is } \{\boldsymbol{D}, \boldsymbol{\beta}^k\}, \text{ with a rule weight } \theta_k \text{ and the attribute weight } \boldsymbol{\delta} \tag{1}$$

where $\boldsymbol{U}$ represents the antecedent attribute vector $(U_1, \ldots, U_T)$, V is the consequent attribute, $\boldsymbol{A}^k$ the packet referential values $\left\{A_1^k, \ldots, A_T^k\right\}$ of antecedents, here A_i^k $(i = 1, \ldots, T)$ is the referential value of the antecedent attribute U_i in the kth rule. T is the total number of antecedent attributes used in the rule; ($\boldsymbol{D}$, $\boldsymbol{\beta}^k$) represents $\boldsymbol{D}$ with belief degrees $\boldsymbol{\beta}^k$, i.e., $\{(D_1, \beta_{1k}), \ldots, (D_N, \beta_{Nk})\}$, where $\boldsymbol{D}$ is the consequent vector $(D_1, \ldots, D_N)$, and $\boldsymbol{\beta}^k$ the vector of the belief degrees $(\beta_{1k}, \ldots, \beta_{Nk})$ for $k \in \{1, \ldots, L\}$, and β_{sk} $(s \in \{1, \ldots, N\})$ represents the belief degree to which D_s is believed to be the consequent if in the kth packet rule the input satisfies the packet antecedents $\boldsymbol{A}^k$. θ_k $(\in \Re^+, k = 1, \ldots, L)$ is the relative weight of the kth rule and $\boldsymbol{\delta}$ is the relative weight vector of the antecedent attributes, $\delta_i \in \Re^+$; $i = 1, \ldots, T$. L (>0) is the number of all the packet rules in the rule base. Moreover, $\sum_{s=1}^{N} \beta_{sk} \leqslant 1$. If $\sum_{s=1}^{N} \beta_{sk} = 1$, the kth packet rule is said to be complete; otherwise, it is incomplete.

In a rule base, a referential set can be a set of meaningful and distinctive evaluation standards for describing an attribute, which are commonly described by linguistic terms to reflect and model the vagueness or imprecision in the concepts. In (2) and (3) illustrated the referential values used in antecedent attributes in a belief rule relevant to safety analysis.

Take for example the following belief rule in safety analysis [17]:

R_k: IF the **failure rate** is *frequent* AND the **consequence severity** is *critical* AND the **failure consequence probability** is *unlikely*, THEN the **safety estimate** is $\{(good, 0), (average, 0), (fair, 0.7), (poor, 0.3)\}$ (2)

This is a special case of the rule in (1) while $N = 4$, $T = 3$, $\boldsymbol{U} = (U_1, U_2, U_3)$ = (failure rate, consequence severity, failure consequence probability), $\boldsymbol{A}^k = \left(A_1^k, A_2^k, A_3^k\right) = (frequent, critical, unlikely)$, $\boldsymbol{D} = (D_1, \ldots, D_N) = (good, average, fair, poor)$, and $\boldsymbol{\beta}^k = (\beta_{1k}, \beta_{2k}, \beta_{3k}, \beta_{4k}) = (0, 0, 0.7, 0.3)$. The consequent V (= safety estimate) is described as a belief distribution representation, stating that it is 70% sure that safety level is *fair* and 30% sure that safety level is *poor*. This kind of rule reflects another kind of uncertainty caused because sometimes evidence available is not sufficient or experts are not 100% certain to believe in a hypothesis but only to degrees of belief. If $\beta_{4k} = 0.2$, then $\sum_{s=1}^{4} \beta_{sk} = 0.9 \leqslant 1$, which means the correlation assessment is incomplete with 10% ignorance, which may be due to lack of knowledge, so this may reflect the incompleteness.

In this paper, to facilitate the more general application cases to handle simultaneously vagueness, incompleteness and uncertainty and more flexible and simpler rule base generation scheme, the belief rule in (1) is extended with belief degrees embedded in the entire possible antecedent terms of each rule as well, for example, the belief rule (2) can be extended as follows:

R_k: IF the **failure rate** is $\{(very\ low, 0), (low, 0), (reasonably\ low, 0), (average, 0), (reasonably\ frequent, 0), (frequent, 0), (highly\ frequent, 1)\}$ AND the **consequence severity** is $\{(negligible, 0), (marginal, 0), (moderate, 0), (critical, 0.3), (catastrophic, 0.7)\}$ AND the **failure consequence probability** is $\{(highly\ unlikely, 0.7), (unlikely, 0.2), (reasonably\ unlikely, 0.1), (likely, 0), (reasonably\ likely, 0), (highly\ likely, 0), (definite, 0)\}$ THEN the **safety estimate** is $\{(Good, 0), (Average, 0.1), (Fair, 0.3), (Poor, 0.6)\}$ (3)