Knowledge-Based Systems 000 (2017) 1–21

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A joint optimization method on parameter and structure for belief-rule-based systems

Long-Hao Yang [a,b], Ying-Ming Wang [a,*], Jun Liu [c], Luis Martínez [b]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian, PR China
[b] Department of Computer Science, University of Jaén, Jaén, Spain
[c] School of Computing and Mathematics, Ulster University, Northern Ireland, UK

ARTICLE INFO

Article history:
Received 5 April 2017
Revised 1 October 2017
Accepted 29 November 2017
Available online xxx

Keywords:
Belief-rule-based system
Parameter optimization
Structure optimization
Generalization error
Heuristic strategy

ABSTRACT

The belief-rule-based system (BRBS) is one of the most visible and fastest growing branches of decision support systems. As the knowledge base in the BRBS, the belief-rule-base (BRB) is required to be equipped with the optimal parameters and structure, which means the optimal value and number of parameters, respectively. Several optimization methods were therefore proposed in the past decade. However, these methods presented different limitations, such as the use of the incomplete parameter optimization model, lack of structure optimization, and so on. Moreover, it is impracticable to determine the optimal parameters and structure of a BRB using the training error because of over-fitting. The present work is focused on the joint optimization on parameter and structure for the BRB. Firstly, a simple example is utilized to illustrate and analyze the generalization capability of the BRBS under different numbers of rules, which unveils the underlying information that the BRBS with a small training error may not have superior approximation performances. Furthermore, by using the Hoeffding inequality theorem in probability theory, it is a constructive proof that the generalization error could be a better choice of criterion and measurement to determine the optimal parameters and structure of a BRB. Based on the above results, a heuristic strategy to optimize the structure of the BRB is proposed, which is followed by a parameter optimization method using the differential evolution (DE) algorithm. Finally, a joint optimization method is introduced to optimize the parameters and structure of the BRB simultaneously. In order to verify the generality and effectiveness of the proposed method, two practical case studies, namely oil pipeline leak detection and bridge risk assessment, are examined to demonstrate how the proposed method can be implemented in the BRB under disjunctive and conjunctive assumptions along with their performance comparative analysis.

1. Introduction

The decision support system is an important tool to address decision problems and thus many kinds of decision support systems have been constructed to deal with the product customization [1–3], medical diagnosis [4,5], and others. Among these decision support systems, the rule-based system has become one of the most useful approaches to model and analyze decision problems using various types of knowledge [6]. Traditional rule-based systems usually adopt simple IF-THEN rules to represent knowledge, e.g. in the form of "*IF presence of creatinine THEN renal failure is definite*" [7]. The rule means that the consequent "*renal failure is definite*" is believed to be true with the probability of 100% from the given cause "*presence of creatinine*". However, previous studies [8–10] have shown that such strict knowledge representation scheme is inefficient in expressing information with uncertainty, such as how to represent the consequent with the probability of 10%, 20%, or 50%. Therefore, there is a need to enhance the knowledge representation scheme in rule-based systems.

Due to the limitations of the traditional IF-THEN rules, many attempts [11,12] have been made to combine other theories, such as the interval probability theory and fuzzy set theory, to improve the capability of expressing uncertain information using IF-THEN rules. The prevailing one is the belief rule, which is a more advanced IF-THEN rule proposed by Yang et al. [13]. One of distinctive improvements is to replace the single-valued consequent by using a distributed assessment called a belief structure. The use of belief structure has enabled various kinds of uncertain information, including probability, fuzzy, and incompleteness. By grouping belief rules into one belief-rule-base (BRB), the belief-rule-based system (BRBS) has an effective knowledge base to address different deci-

* Corresponding author.
E-mail address: ymwang@fzu.edu.cn (Y.-M. Wang).

https://doi.org/10.1016/j.knosys.2017.11.039
0950-7051/

sion problems. To date, the applications of the BRBS include different areas, such as the prediction of consumer preferences [14–16], detection of pipeline leak [17,18], prediction of Forex trading [19], Assessment of product lives [20], safe evaluation of engineering systems [21–23], clinical decision support [24–26] and basic classification problems [27–29].

However, two basic issues for setting up a BRB should be solved before the corresponding BRBS is ready for an application.

(1) Determination of the optimal parameters involved in the BRB, which means to optimize the value of parameters for a BRB, such as the antecedent attribute weights, the rule weights, the utility values of the referential values used for each antecedent attribute, and the belief degrees attached to the consequent. The optimal parameters are closely related to the performance of the established BRBS. Accordingly, many parameter optimization models and techniques have been developed, such as the local parameter optimization model [30], the adaptive parameter optimization model [31], and the swarm intelligent algorithm-based optimization technique [26,32], etc.
(2) Determination of the optimal structure involved in the BRB, which means to optimize the number of parameters for a BRB. The optimal structure is closely linked to the size of BRB and affects the performance of the BRBS in terms of both accuracy and efficiency. The big number may lead to an over-fitting, whilst too small number may result in an under-fitting. Thus, many structure optimization approaches, including the sequential learning algorithm [31], dimensionality reduction-based structure learning method [33], the dynamic rule adjustment approach [34], etc, have been proposed in order to obtain an optimal structure of the BRB.

Based on the literature review for solving the above two basic issues, several deficiencies could be found. On the one hand, some attempts may not be able to produce promising results in optimizing a BRB because of their limitations. For instance, the local parameter optimization model [30] cannot achieve better accuracy than the adaptive parameter optimization model [31] under the same experimental conditions. On the other hand, some attempts are only available to determine the optimal parameters [30,31] without the optimal structure of a BRB. It is evident that both are important and necessary for constructing an optimal BRB [35]. More importantly, for most of attempts to optimize the parameters and the structure of a BRB, such as the sequential learning algorithm [35] and the dynamic rule adjustment approach [34], the minimal training error obtained by optimizing the parameters is regarded as an important criterion to determine the optimal structure. However, the minimal training error might be derived from the BRBS with the over-fitting.

To overcome the above limitations, the following challenges need to be addressed:

(1) Identifying a new criterion to determine the optimal parameters and structure of a BRB to avoid the over-fitting.
(2) Designing an efficient strategy to obtain the optimal one from the large number of possible the structure of BRBs.
(3) Developing a generic optimization technique to obtain the optimal parameters of a BRB from arbitrary a structure.

In the present work, to address the first challenge, a simple example is used to analyze and illustrate relationship between the generalization capability of BRBS and the size of BRB. Thus, three specific BRBs with different numbers of rules are designed to reveal the relationship between the approximation error and the desired number of rules. Furthermore, by using the Hoeffding inequality theorem [36] to link the training error, the number of training data, and the structure of the BRB, the concept of generalization error is proposed as a new criterion to determine the optimal parameters and structure of a BRB.

For the second challenge, the large number of possible BRBs options force that an efficient strategy is required to be proposed to obtain the optimal structure of a BRB. For example, based on the arbitrary number of referential values that can construct different BRBs with different structures, the optimal structure of the BRB has to be obtained from $2^6 = 64$, $2^8 = 256$ and $2^{10} = 1024$ possible BRBs options if with two antecedent attributes and each attribute used three, four and five referential values respectively. In the present work, the set of referential values is used as a concise scheme to express the structure of a BRB. Then a heuristic strategy based on the generalization error is proposed to efficiently optimize the number of referential values for each antecedent attribute among all the possible BRBs.

For the third challenge, although different parameter optimization techniques of the BRB have been developed, three dilemmas must be addressed: (1) the stop criterion has a great impact on optimization results; (2) different parameter optimization models have different constraints for the same parameter, some of which may be dynamic; (3) the referential values are normally expressed as linguistic terms, so there should be an equivalent transformation between the utility values of the referential values and the corresponding linguistic terms. Accordingly, in the present work, the differential evolutionary (DE) algorithm [37] is adopted as a generic technique to address the above dilemmas in order to obtain the optimal parameters of a BRB under arbitrary a structure.

On the basis of the proposed solutions for the above three challenges, a new joint optimization method is proposed to obtain both the optimal parameters and the optimal structure of the BRB simultaneously. The key idea is summarized as follows: the above mentioned heuristic strategy and the DE algorithm-based technique are responsible for optimizing the structure and the parameters of the BRB, respectively. The generalization error is then used to judge whether the BRB is superior to other BRBs in terms of the structure and parameters. Finally, the BRB with the minimal generalization error is selected as the optimal one.

To illustrate the generality and effectiveness of the proposed method, the widely used case study on oil pipeline leak detection [17,18] and bridge risk assessment [38,39] are conducted on the BRB under conjunctive and disjunctive assumptions. Two main aspects, namely, accuracy and the number of parameters, are used to compare with other BRBSs and the corresponding parameter and structure optimization methods.

The rest of the paper is organized as follows: Section 2 briefly reviews the basics of the BRBS and its related works. Section 3 introduces and illustrates the concept of generalization error. Section 4 proposes the joint optimization method for obtaining the optimal parameters and the optimal structure of a BRB. Section 5 provides two case studies to demonstrate the generality and effectiveness of the proposed method, and the paper is concluded in Section 6.

2. Brief introduction to BRBS

2.1. Basics of BRBS

A BRBS consists of two main components. The first is the BRB that can be regarded as the knowledge base to save various kinds of uncertain information. The second is the evidential reasoning (ER)-based inference method that provides an inference engine to reply input data according to the BRB. A simple methodological framework of the BRBS is shown in Fig. 1.

For a BRB, suppose it has M antecedent attributes and one consequent attribute, each antecedent attribute U_i ($i = 1,\ldots,M$) has J_i referential values $A_{i,j}$ which are used for describing the ith an-