ARTICLE IN PRESS

Information Sciences xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

On the inference and approximation properties of belief rule based systems

Yu-Wang Chen [a,*], Jian-Bo Yang [a,b], Dong-Ling Xu [a,b], Shan-Lin Yang [b]

[a] Decision and Cognitive Sciences Research Centre, Manchester Business School, The University of Manchester, Manchester M15 6PB, United Kingdom
[b] School of Management, Hefei University of Technology, Hefei, Anhui 230009, China

ARTICLE INFO

Article history:
Received 21 April 2011
Received in revised form 15 January 2013
Accepted 22 January 2013
Available online xxxx

Keywords:
Belief rule base
Evidential reasoning
Inference
Universal approximation
System identification

ABSTRACT

Belief rule based (BRB) system provides a generic inference framework for approximating complicated nonlinear causal relationships between antecedent inputs and output. It has been successfully applied to a wide range of areas, such as fault diagnosis, system identification and decision analysis. In this paper, we provide analytical and theoretical analyses on the inference and approximation properties of BRB systems. We first investigate the unified multi-model decomposition structure of BRB systems, under which the input space is partitioned into different local regions. Then we analyse the distributed approximation process of BRB systems. These analysis results unveil the underlying inference mechanisms that enable BRB systems to have superior approximation performances. Furthermore, by using the Stone–Weierstrass theorem, we constructively prove that BRB systems can approximate any continuous function on a compact set with arbitrary accuracy. This result provides a theoretical foundation for using and training BRB systems in practical applications. Finally, a numerical simulation study on the well-known benchmark nonlinear system identification problem of Box–Jenkins gas furnace is conducted to illustrate the validity of a BRB system and show its inference and approximation capability.

1. Introduction

Belief rule based (BRB) system, as an extension of traditional IF–THEN rule based system, has been successfully applied in various areas, such as fault diagnosis, system identification and decision analysis [1,18,27,31,32,35,36]. When applying a BRB system, the input of each antecedent attribute is transformed into a belief distribution over a set of referential values. The belief distribution is further used to calculate the activation weights of belief rules in the belief rule base. Subsequently, the inference output is generated through the aggregation of all activated belief rules using the evidential reasoning (ER) approach [28,30]. This methodology is developed on the basis of the Dempster-Shafer theory of evidence [6,20], the decision theory [9,33], rule-based systems [12,17] and relevant artificial intelligence and information techniques [8,16,22]. Compared with traditional rule based systems, BRB systems provide a more informative knowledge representation scheme for both quantitative data and qualitative information with uncertainties, and it is capable of approximating complicated nonlinear causal relationships between antecedent inputs and output [31]. In a BRB system, belief rule base is one of the most important components, in which belief rules can be established initially by human experts with domain-specific knowledge and facts. However, by using experts' subjective knowledge alone, it is difficult to determine the parameters of a belief rule base

* Corresponding author.
E-mail address: yu-wang.chen@mbs.ac.uk (Y.-W. Chen).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.01.022

accurately, in particular for practical rule bases with many rules. Consequently, optimal learning models and relevant optimisation techniques have also been proposed for training the parameters of BRB systems [5,32,35].

The main purpose to construct a BRB system is to approximate a real system as accurately as possible, and also to avoid overfitting the training data. The following two fundamental theoretical questions on the inference and approximation properties of BRB systems therefore need to be studied thoroughly for supporting practical applications.

(1) What are the underlying inference mechanisms which enable BRB systems to have the superior approximation performances as claimed in the literatures in a range of applications?
(2) Are BRB systems universal approximators? In other words, are BRB systems capable of approximating any continuous function on a compact set with arbitrary accuracy?

Similar theoretical questions were investigated in the development of artificial neural networks (ANN) and fuzzy systems [3,11,13,23,24,34]. In this paper we investigate these issues to underpin the theoretical foundation of using and training BRB systems in practical applications. First of all, we analyse the multi-model decomposition structure and the distributed approximation process of BRB systems. By utilising the Stone–Weierstrass theorem [4,21], we then constructively prove that BRB systems can approximate any continuous function on a compact set with arbitrary accuracy. Finally, a numerical simulation study on the well-known benchmark nonlinear system identification problem of Box–Jenkins gas furnace [2] is conducted to demonstrate the validity of a BRB system and show its inference and approximation capability.

The rest of this paper is organised as follows. In the following section, the inference process of BRB systems is briefly discussed. In Section 3, a unified multi-model decomposition structure for partitioning input space into different local regions is investigated and the distributed approximation in the process of inferring output is analysed. Based on the Stone–Weierstrass theorem, the universal approximation property of BRB systems is proved in Section 4. In Section 5, a numerical simulation study is conducted on the nonlinear system identification problem of Box–Jenkins gas furnace. The paper is concluded and comparisons with some state-of-the-art methodologies and further research are discussed in Section 6.

2. Description of belief rule based systems

In this paper we consider a numerical multi-input single-output system f: $U \subset R^M \to R$, whilst a multi-output system can always be separated into a set of single-output systems [10,24].

2.1. Belief rule base

For a multi-input system, suppose the following M antecedent attributes are a complete set of factors influencing the system's output, i.e.,

$$X = \{x_i, i = 1, \ldots, M\} \tag{1}$$

In order to represent the behaviours of the multi-input system, a finite number of belief rules can be constructed, which constitute a belief rule base. Formally, a belief rule is defined as follows:

$$\begin{aligned} R_k: \ & \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_{M_k} \text{ is } A_{M_k}^k \\ & \text{THEN } \{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}), \ldots, (D_N, \beta_{N,k})\}, \quad \left(\sum_{n=1}^{N} \beta_{n,k} \leqslant 1\right) \\ & \text{with rule weight } \theta_k \\ & \text{and attribute weight } \delta_{1,k}, \delta_{2,k}, \ldots, \delta_{M_k,k}, \quad k \in \{1, \ldots, K\} \end{aligned} \tag{2}$$

where $x_1, x_2, \ldots, x_{M_k}$ denote the antecedent attributes in the kth rule, and these attributes can be a subset of all the antecedent attributes $X = \{x_i; i = 1, \ldots, M\}$. $A_i^k (i = 1, \ldots, M_k)$ is the referential value taken by the ith antecedent attribute in the kth rule and $A_i^k \in A_i$. $A_i = \{A_{i,j}; j = 1, \ldots, J_i\}$ denotes the set of referential values for the ith antecedent attribute and J_i is the number of the referential values. In BRB systems, the referential values can also be labelled as linguistic terms, such as "small", "medium", "large" and so forth. $\beta_{n,k}$ $(n = 1, \ldots, N; k = 1, \ldots, K)$ represents the belief degree to which the consequent element D_n is believed to be the consequent, given the logical relationship of the kth rule IF x_1 is $A_1^k \wedge x_2$ is $A_2^k \wedge \cdots \wedge x_{M_k}$ is $A_{M_k}^k$. Note that the element D_n in the set of consequent elements $D = \{D_n; n = 1, \ldots, N\}$ can either be a conclusion or an action and a subset of elements can also be part of the consequent [26]. If $\sum_{n=1}^{N} \beta_{n,k} = 1$, the kth rule is said to be complete; otherwise, it is incomplete. θ_k is the relative weight of the kth rule, and $\delta_{M_k,k}$ represents the relative weight of attributes in the kth rule.

Under the unified scheme as defined above, a belief rule can represent a functional mapping between antecedent inputs and output possibly with uncertainties. It provides a more informative and realistic scheme than a traditional IF–THEN rule. Here, it is worth noting that human knowledge representation forms can always be transformed into IF–THEN rule-based schemes [22]. Once a belief rule base is established, the knowledge embedded in all belief rules can be used to perform inference for a specific input vector.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

2.2. Inference using the Evidential Reasoning (ER) approach

Suppose we have an input vector $x(t) = \{x_i(t), i = 1, \ldots, M\}$ at the sampling time t in a complex nonlinear system. For simplicity, x_i is used to represent $x_i(t)$ afterwards. Using the referential values defined for the ith antecedent attribute x_i, the input x_i can be transformed to the following belief distribution [29].

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}), j = 1, \ldots, J_i\} \tag{3}$$

where

$$\alpha_{i,j} = \frac{A_{i,j+1} - x_i}{A_{i,j+1} - A_{i,j}} \quad \text{and } \alpha_{i,j+1} = 1 - \alpha_{i,j}, \quad \text{if } A_{i,j} \leqslant x_i \leqslant A_{i,j+1}$$
$$\alpha_{i,j'} = 0, \quad \text{for } j' = 1, \ldots, J_i \quad \text{and } j' \neq j, j+1$$

Here, $\alpha_{i,j}$ represents the similarity degree to which the input value x_i matches the referential value $A_{i,j}$. After all the inputs are transformed into belief distributions, the activation weight of the kth belief rule can be calculated as follows [31].

$$w_k(x) = \frac{\theta_k \prod_{i=1}^{M_k} \left(\alpha_{i,j}^k\right)^{\bar{\delta}_i}}{\sum_{l=1}^{K} \left[\theta_l \prod_{i=1}^{M_l} (\alpha_{i,j}^l)^{\bar{\delta}_i}\right]}, \quad \text{and } \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M_k}\{\delta_i\}} \tag{4}$$

Note that in Eq. (4) the attribute weights are assumed to be identical in all belief rules. Further, the belief degrees on the inference output can be generated through the aggregation of all activated belief rules using the following analytical ER approach [25].

$$\beta_n(x) = \frac{\prod_{k=1}^{K}\left(w_k(x)\beta_{n,k} + 1 - w_k(x)\sum_{i=1}^{N}\beta_{i,k}\right) - \prod_{k=1}^{K}\left(1 - w_k(x)\sum_{i=1}^{N}\beta_{i,k}\right)}{\sum_{j=1}^{N}\prod_{k=1}^{K}\left(w_k(x)\beta_{j,k} + 1 - w_k(x)\sum_{i=1}^{N}\beta_{i,k}\right) - (N-1)\prod_{k=1}^{K}\left(1 - w_k(x)\sum_{i=1}^{N}\beta_{i,k}\right) - \prod_{k=1}^{K}(1 - w_k(x))} \tag{5}$$

As a result, the inference output can be represented as the following belief distribution.

$$S(f(x)) = \{(D_n, \beta_n(x)), n = 1, \ldots, N\} \tag{6}$$

If we know the utility $u(D_n)$ for each consequent element D_n, the numerical inference output can be calculated by

$$f(x) = \sum_{n=1}^{N} u(D_n)\beta_n(x) \tag{7}$$

It is worth noting that the utility $u(D_n)$ is also adjustable according to experts' preferences and/or observed output data.

In summary, a BRB system has the following design parameters:

(1) The number of rules K, which is equal to $\prod_{i=1}^{M} J_i$ for a complete belief rule base including all the possible combinations of referential values for all antecedent attributes;
(2) Adjustable parameters, including referential value $A_{i,j}$, the number of referential values J_i, belief degree $\beta_{n,k}$, the utility $u(D_n)$ of each consequent element D_n, the number of consequent elements N, rule weight θ_k and attribute weight δ_i.

Note that all the design parameters in a BRB system can be associated with physical meanings in specific applications. For example, the number of referential values for antecedent attributes determines the dimensionality of a complete rule base. These design parameters provide BRB systems with a high degree of flexibility. The values of these parameters can be assigned initially by domain experts. However, if observed input–output data pairs $(x(t), \hat{f}(x(t)))$ are available, optimal learning methods can be designed to train the parameters for minimising the difference between the inference output of a BRB system and the observed output. The following mean squared error (MSE) can be used to measure the difference between inference output and observed output.

$$\xi(P) = \frac{1}{T}\sum_{t=1}^{T}(f(x(t)) - \hat{f}(x(t)))^2 \tag{8}$$

where $P = \langle A_{i,j}, J_i, \beta_{n,k}, \theta_k, \delta_i, u(D_n), N\rangle$ denotes the vector of design parameters. According to their physical meanings and functional requirements, all parameters must satisfy certain equality and inequality constraints [5,32].

3. Inference mechanisms of BRB systems

In terms of mathematical modelling, a BRB inference system formulates a nonlinear functional mapping between antecedent input space and output. In this section, the kernel of the modelling and inference mechanisms of a BRB system is analysed both analytically and graphically.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

3.1. Multi-model decomposition of input space

To construct a BRB system, a number of referential values need to be defined first for each antecedent attribute, and the antecedents in a belief rule are the combination of the referential values of antecedent attributes. As such, the input space U can be decomposed into multiple local regions by the referential values, which can be represented by the hyperspace $[A_{1,1}, A_{1,J_1}] \times \cdots \times [A_{M,1}, A_{1,J_M}]$. Take a two-dimensional input space for example. The referential values can decompose the input space $x_1 \times x_2$ into $(J_1 - 1)(J_2 - 1)$ rectangular local regions as shown in Fig. 1.

The number of referential values on antecedent attributes decides the granularity and interpretability of the local regions. Generally speaking, antecedent inputs which result in a high degree of variation on output require more referential values than otherwise. The intersection points of adjacent local regions are the belief rule points at which the combinations of referential values are the antecedents of belief rules. If a numerical input vector $x(t)$ falls into a specific local region formed by a set of belief rule points, as shown by the solid points in Fig. 1, these belief rules will be activated to infer the estimated output $f(x(t))$.

Under this structure, a BRB system is in essence a multi-model approximator in which the local regions represent decomposed sub-models that are combined to describe the global behaviour of the desired system. For each local model, the relationship between antecedent inputs and output is formulated by the unified inference scheme as defined by Eqs. (3)–(5), (3), (7). To demonstrate the local inference and global approximation capability of BRB systems, the inference mechanisms need to be studied thoroughly. In a numerical BRB system, we have $\sum_{n=1}^{N} \beta_{n,k} = 1$, and then Eq. (5) can be transformed as follows:

$$\beta_n(x) = \frac{\prod_{k=1}^{K}(w_k(x)\beta_{n,k} + 1 - w_k(x)) - \prod_{k=1}^{K}(1 - w_k(x))}{\sum_{j=1}^{N}\left[\prod_{k=1}^{K}(w_k(x)\beta_{j,k} + 1 - w_k(x)) - \prod_{k=1}^{K}(1 - w_k(x))\right]} \tag{9}$$

Due to the recursive nature of the ER approach, we only need to consider two typical belief rules R_k and R_l in the following theoretical analysis, and the belief rule R_k is assumed to be an intermediate belief rule generated by aggregating all other activated belief rules except for R_l. Suppose the belief distributions of the two belief rules R_k and R_l are given as follows:

$$R_k : \{(D_n, \beta_{n,k}), n = 1, \ldots, N\} \quad \text{with} \sum_{n=1}^{N} \beta_{n,k} = 1$$

$$R_l : \{(D_n, \beta_{n,l}), n = 1, \ldots, N\} \quad \text{with} \sum_{n=1}^{N} \beta_{n,l} = 1$$

In what follows, we study the properties of BRB systems, including continuity, boundedness, derivability and monotonicity.

Lemma 1. Continuity property:*The output of a BRB system is continuous.*

Proof. In a BRB system, each input falls into a specific local region or on the boundary of adjacent local regions. According to the analytical reasoning model given in Eqs. (3)–(7), it is obvious that the inference output of a BRB system is continuous within a local region. As a result, we only need to prove the continuity at the intersection rule points.

To calculate the limit of $\beta_n(x)$ at the intersection rule point R_l from any direction, assume $w_l(x) = w(x)$, and so $w_k(x) = 1 - w(x)$, given that R_k is an intermediate belief rule aggregated from all activated belief rules adjacent to R_l. We then have

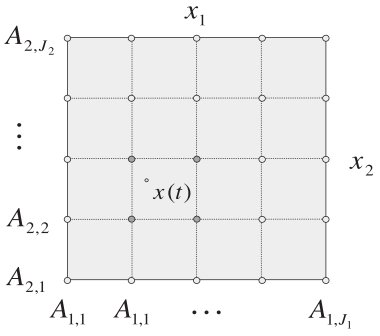

Fig. 1. Input space decomposition and belief rule points.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

$$\beta_n(x) = \frac{(\beta_{n,l} + \beta_{n,k} - \beta_{n,l}\beta_{n,k})w^2(x) + (\beta_{n,l}\beta_{n,k} - 2\beta_{n,k})w(x) + \beta_{n,k}}{\left(2 - \sum_{j=1}^{N}(\beta_{j,l}\beta_{j,k})\right)w^2(x) - \left(2 - \sum_{j=1}^{N}(\beta_{j,l}\beta_{j,k})\right)w(x) + 1} \tag{10}$$

According to Eq. (4), we have $w(x) \to 1^-$ if the input $x(t)$ approaches the antecedents of the intersection rule point R_l infinitely from any direction. Furthermore, from Eq. (10) we then have

$$\lim_{w(x)\to 1^-} \beta_n(x) = \beta_{n,l} \tag{11}$$

The above result shows that the limits of the belief degrees in the consequent of the estimated output $f(x)$ are independent of any activated belief rule adjacent to R_l (i.e. R_k) while the input $x(t)$ approaches the antecedents of R_l infinitely. A BRB system is thus continuous at any intersection rule point. Therefore, we prove that a BRB system is a continuous system.

Corollary 1. Uniform continuity property:*The output of a BRB system is uniformly continuous.*

Proof. It has been proved in Lemma 1 that a BRB system is a continuous system. In addition, we know that the input space of a BRB system is actually defined on a compact metric space $[A_{1,1}, A_{1,J_i}] \times \cdots \times [A_{M,1}, A_{1,J_M}]$. According to the Heine-Cantor theorem in mathematics, a continuous function on a compact set is uniformly continuous [37]. Thus we prove that a BRB system is a uniformly continuous system.

Lemma 2 (*Boundedness property*). *The output of a BRB system is always bounded.*

Proof. According to the analytical ER algorithm in Eq. (5), we have

$$\beta_n(x) \leqslant 1, \forall x(t) \in U, \quad \text{and } f(x) = \sum_{n=1}^{N} u(D_n)\beta_n(x)$$

As discussed above, we also define a set of bounded utilities to measure the individual consequent element D_n. Without loss of generality, we suppose that D_1 is the least preferred consequent element having the lowest utility and D_N is the most preferred consequent element having the highest utility, i.e., $u(D_1) < \cdots < u(D_N)$. Thus we have $u(D_1) \leqslant f(x) \leqslant u(D_N)$, and we prove that a BRB system is a bounded system.

Lemma 3. *The output of a BRB system is not necessarily derivable.*

Proof. We can calculate the partial first order-derivatives of a BRB system at the intersection rule point R_l as follows:

$$\left.\frac{\partial \beta_n(x)}{\partial w(x)}\right|_{w(x)=-1^-} = \beta_{n,l}\left(\sum_{j=1}^{N}(\beta_{j,l}\beta_{j,k}) - \beta_{n,k}\right) \tag{12}$$

It shows that the first order-derivatives of a BRB system at the intersection rule point R_l are dependent on the parameters of both the belief rule R_l and the intermediate aggregated belief rule R_k. Since different belief rules may be activated when the input approaches the intersection rule point R_l from different directions, there is no guarantee that the partial first-order derivatives at the intersection rule point R_l from all directions would be equal. This means that the output of a BRB system is not necessarily derivable at intersection rule points, although it is obvious from Eqs. (3)–(7) that the output is derivable within a local region. Thus we can conclude that a BRB system is not necessarily a derivable system.

Remark 1. Although the output of a general BRB system is not necessarily derivable and provides no information about the gradients of the approximated hyper-surfaces at intersection rule points, an application-specific BRB system can be designed to be derivable if constraints are added to ensure that the first order-derivatives from all directions are equal at any intersection rule point.

Lemma 4. *The output of a BRB system is not necessarily monotonic.*

Proof. Without loss of generality, we consider a single input system. It is obvious from Eq. (4) that the activation weight $w(x)$ of the belief rule R_l is monotonic with respect to the input value x which falls between the two referential values of R_k and R_l. Further, the first-order derivative of the inference output $f(x)$ with respect to the activation weight $w(x)$ can be calculated by

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

$$\frac{\partial f(x)}{\partial w(x)} = \sum_{n=1}^{N} u_n \frac{\partial \beta_n(x)}{\partial w(x)} \tag{13}$$

Since it was proven that $\frac{\partial \beta_n(x)}{\partial w(x)}$ is not necessarily nonnegative [5], their linear combination is not necessarily nonnegative either. This means that the output of a BRB system is not necessarily monotonic.

According to Lemmas 1–4, we can conclude that a BRB system is uniformly continuous and always bounded but not necessarily derivable or monotonic.

3.2. Distributed approximation process of BRB systems

A BRB system is a distributed approximation process, with its inference output represented by a belief distribution $\{(D_n, \beta_n(x)), n = 1, \ldots, N\}$. Eq. (7) shows that the numerical output of a BRB system is the linear combination of the products of the combined belief degree $\beta_n(x)$ and the utility $u(D_n)$. Utility can be defined using experts' subjective judgements and/or observed output data. The combined belief degree $\beta_n(x)$ is generated by aggregating the belief distributions of consequents in all activated rules. In the following, we use two simple examples to illustrate and explain the BRB distributed approximation process.

Example 1. Suppose there is a single-input system, whose input x is bounded in the interval [0, 1] and whose output $g(x)$ is a monotonically increasing function of x and is bounded by $g(0) = 0$ and $g(1) = 1$. To construct a BRB system, the referential utilities of five consequent elements in belief rules are defined as follows:

$$\{D_1, D_2, D_3, D_4, D_5\} = \{-0.5, 0, 0.5, 1, 1.5\}$$

We demonstrate that a BRB system with only two belief rules can have remarkable approximation power to simulate a monotonically increasing system that can be linear, convex or concave with high degrees of accuracy. An initial BRB system is shown in Table 1.

To illustrate the distributed approximation process in the interval [0, 1], we use a collection of power functions, i.e. $g(x) = x^{\tau}$ with the exponent parameter τ set to 6, 4, 2, 1, 1/2, 1/4 and 1/6 respectively, to generate observed input–output data pairs. For each τ, a set of 101 data pairs in the interval is generated uniformly. The data sets are used to train the parameters in the BRB system shown in Table 1, and the 7 trained approximation curves are shown in Fig. 2.

In Fig. 2, the dotted line is the output of the initial BRB system shown in Table 1, and the solid lines from bottom are the approximation curves from 1 to 7 respectively. The MSEs calculated for the seven sets of data pairs are 2.95×10^{-4}, 2.55×10^{-4}, 9.70×10^{-5}, 1.62×10^{-7}, 1.27×10^{-4}, 3.48×10^{-4}, and 3.14×10^{-4} respectively. Note that the MSEs can be reduced further if more consequent elements are used. It is evident from the figure and the MSEs that the BRB system has superior approximation capability through adjusting its design parameters. Table 2 lists the parameter values of the seven trained BRB systems.

It is clear from Eq. (4) that rule weights θ_k and attribute weights δ_i have direct influences on activation weights. One can see in Table 2 that the attribute weight is changed considerably, updated to be monotonically decreasing for generating the seven approximation curves from convex, linear to concave as shown in Fig. 2. From the convex curves 1 to 3, the weight of rule 2 is updated to be monotonically increasing, with the weight of rule 1 kept almost constant; from the concave curves 5 to 7, the weight of rule 1 is updated to be monotonically decreasing, with the weight of rule 2 kept almost constant; for the line 4, the weights of both rules 1 and 2 are equal. These simulation results show that the change of rule weights has significant effects on the approximation patterns. For example, curve 1 has a high degree of convexity, and the output is closer to the consequent of rule 1 in a large range of the input interval [0, 1]. Thus the updated weight of rule 2 is small, and rule 1 plays a more important role in the aggregation of the inference output. On the other hand, curve 7 has a high degree of concavity, and the output is closer to the consequent of rule 2 in a large range of the input interval [0, 1]. Thus the updated weight of rule 1 is small, and rule 2 plays a more important role in the aggregation of inference output.

Also note that the belief distributions of consequents in all trained belief rules are updated as well. In the initial BRB system, only one consequent element in each belief rule is associated with nonzero belief degree. In the trained BRB systems, however, the belief distributions are more widely distributed and most of the consequent elements are associated with nonzero belief degrees. In general, a more widely distributed belief distribution will be generated for aggregating the inference output of a trained BRB system if a system exhibits more complicated patterns. In this example, for instance, suppose we aggregate the desired output $f(x) = 0.5$ from the 7 approximation curves. It is worth noting that the actual inputs for generating the desired output from curves 1 to 7 are 0.8909, 0.8409, 0.7071, 0.5000, 0.2500, 0.0625, and 0.0156,

Table 1
Initial BRB system for Example 1.

Rule no.	Rule weight	x	$f(x)$	Consequents $\{D_1, D_2, D_3, D_4, D_5\} = \{-0.5, 0, 0.5, 1, 1.5\}$
1	1	0	0	$\{(D_1, 0), (D_2, 1), (D_3, 0), (D_4, 0), (D_5, 0)\}$
2	1	1	1	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 1), (D_5, 0)\}$

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

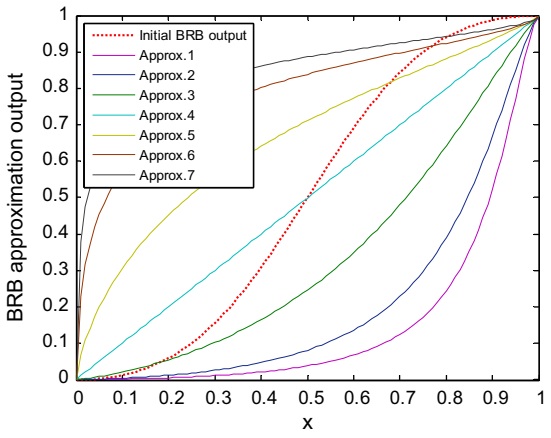

Fig. 2. Illustration of the distributed approximation process of BRB systems.

Table 2
Trained BRB systems for Example 1.

Approx.	Attribute weight	Rule no.	Rule weight	x	$g(x)$	Belief degrees on $\{D_1, D_2, D_3, D_4, D_5\}$
1 ($\tau = 6$)	0.7595	1	0.9847	0	0	(0.0056, 0.9918, 0.0005, 0.0011, 0.0010)
		2	0.1948	1	1	(0.0000, 0.0101, 0.0000, 0.9696, 0.0203)
2($\tau = 4$)	0.7208	1	0.9942	0	0	(0.0000, 1.0000, 0.0000, 0.0000, 0.0000)
		2	0.2968	1	1	(0.0000, 0.0385, 0.0052, 0.8742, 0.0821)
3 ($\tau = 2$)	0.6314	1	1.0000	0	0	(0.0080, 0.9893, 0.0000, 0.0000, 0.0027)
		2	0.5806	1	1	(0.0000, 0.0628, 0.0000, 0.8116, 0.1256)
4 ($\tau = 1$)	0.5101	1	1.0000	0	0	(0.0429, 0.9359, 0.0001, 0.0205, 0.0006)
		2	1.0000	1	1	(0.0006, 0.0205, 0.0001, 0.9359, 0.0429)
5 ($\tau = 1/2$)	0.4049	1	0.6476	0	0	(0.1283, 0.8068, 0.0014, 0.0635, 0.0000)
		2	0.9978	1	1	(0.0002, 0.0003, 0.0002, 0.9981, 0.0012)
6 ($\tau = 1/4$)	0.3218	1	0.4111	0	0	(0.0734, 0.8902, 0.0002, 0.0355, 0.0007)
		2	0.9643	1	1	(0.0018, 0.0312, 0.0010, 0.8970, 0.0690)
7 ($\tau = 1/6$)	0.2863	1	0.3371	0	0	(0.0425, 0.9362, 0.0000, 0.0213, 0.0000)
		2	0.9780	1	1	(0.0002, 0.0301, 0.0001, 0.9086, 0.0610)

Table 3
Aggregated belief distributions for representing the desired output $f(x) = 0.5$.

Approx.	Real input	Desired output	Aggregated belief degrees on $\{D_1, D_2, D_3, D_4, D_5\}$	Estimated output
1 ($\tau = 6$)	0.8909	0.5	(0.0029, 0.5157, 0.0003, 0.4709, 0.0104)	0.4852
2 ($\tau = 4$)	0.8409	0.5	(0.0000, 0.5325, 0.0025, 0.4251, 0.0399)	0.4862
3 ($\tau = 2$)	0.7071	0.5	(0.0038, 0.5345, 0.0000, 0.3985, 0.0632)	0.4914
4 ($\tau = 1$)	0.5000	0.5	(0.0214, 0.4786, 0.0000, 0.4786, 0.0214)	0.5000
5 ($\tau = 1/2$)	0.2500	0.5	(0.0631, 0.3961, 0.0008, 0.5395, 0.0006)	0.5091
6 ($\tau = 1/4$)	0.0625	0.5	(0.0372, 0.4686, 0.0006, 0.4604, 0.0332)	0.4918
7 ($\tau = 1/6$)	0.0156	0.5	(0.0233, 0.5389, 0.0000, 0.4115, 0.0263)	0.4393

respectively. In the initial BRB system, the belief distribution (0, 0.5, 0, 0.5, 0) is aggregated to represent the desired output. In the trained BRB system, however, the inference outputs for approximation curves 1–7 are represented by the following more distributed belief distributions as listed in Table 3.

It is obvious that the above seven belief distributions are quite different, although they are aggregated to represent the same desired output $f(x) = 0.5$. This is because in different curves the input and the belief rules used to aggregate the same desired output are different. Furthermore, according to the synthesis axioms of the ER approach [30], if the consequents in all activated belief rules are completely assessed to a subset of consequent elements, then the aggregated consequent should be completely assessed to the same subset as well. This implies that a more distributed belief distribution will be generated by the ER approach if the belief distributions of consequents in all activated belief rules are updated to be more widely distributed in trained BRB systems. It is evident from Fig. 2 that the trained BRB systems with more distributed belief

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

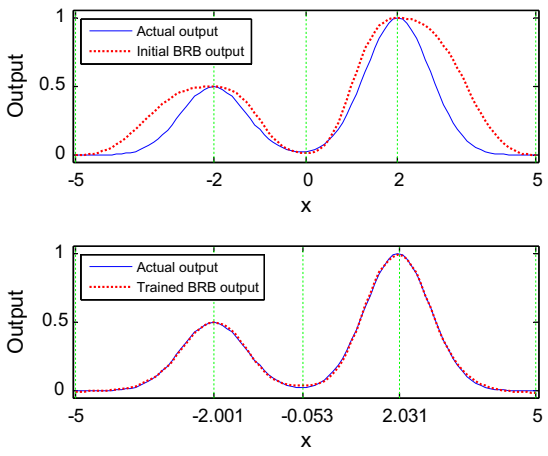

Fig. 3. Illustration on multi-model decomposition and distributed approximation in Example 2.

distributions of consequents in all belief rules have higher approximation flexibility. Thus, it can be concluded that a BRB system is a distributed approximation process, which enables it to have superior approximation capability as illustrated in this example.

Remark 2. One of the strengths of this distributed approximation process is that only a small number of belief rules may be needed to approximate a complex nonlinear system, which may have complicated behaviours but from which only a limited number of data sets may be sampled.

In the following, we use a multi-extremal function to illustrate how a BRB system with a small number of dedicated belief rules can be constructed to approximate a complicated nonlinear system if some behaviours of the system are known *a priori*.

Example 2. A multi-extremal function is given by

$$g(x) = e^{-(x-2)^2} + 0.5e^{-(x+2)^2}, \quad -5 \leqslant x \leqslant 5$$

For this function, we know that approximately $g(x)$ has a local minimum at point $x = 0$, a local maximum at point $x = -2$ and a global maximum at point $x = 2$. Based on the observed outputs on these critical points, the following referential utilities can be defined for the consequents of belief rules.

$$\{D_1, D_2, D_3, D_4, D_5\} = \{-0.5, 0, 0.5, 1, 1.5\}$$

As shown in Fig. 3, $g(x)$ roughly has four monotonic regions in the interval $-5 \leqslant x \leqslant 5$, i.e. $-5 \leqslant x \leqslant -2$, $-2 \leqslant x \leqslant 0$, $0 \leqslant x \leqslant 2$ and $2 \leqslant x \leqslant 5$ with five end points. The analyses as conducted in Section 3.1 and Example 1 suggest that it should be sufficient to use five belief rules to approximate this seemingly complicated nonlinear function. The initial belief rules are listed in Table 4 by calculating the values of $g(x)$ at the five end points of the four approximately monotonic regions.

In Fig. 3, one can see that the initial BRB system, although capable of capturing the main changing patterns of the function, is not accurate to simulate the functional model. We can employ the optimal learning method as discussed in Section 2.2 to train the parameters for minimising the difference between inference output and actual output. 101 data sets are uniformly sampled from the interval $[-5, 5]$. In Fig. 3 the actual output and the inference output of the trained belief rule system are compared.

Table 4
Initial belief rules for Example 2.

Rule no.	Rule weight	x	$g(x)$	Consequents $\{D_1, D_2, D_3, D_4, D_5\}$ = {-0.5, 0, 0.5, 1, 1.5}
1	1	−5	0.0001	{(D_1, 0), (D_2, 0.9999), (D_3, 0.0001), (D_4, 0), (D_5, 0)}
2	1	−2	0.5000	{(D_1, 0), (D_2, 0), (D_3, 1), (D_4, 0), (D_5, 0)}
3	1	0	0.0275	{(D_1, 0), (D_2, 0.9725), (D_3, 0.0275), (D_4, 0), (D_5, 0)}
4	1	2	1.0000	{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 1), (D_5, 0)}
5	1	5	0.0001	{(D_1, 0), (D_2, 0.9998), (D_3, 0.0002), (D_4, 0), (D_5, 0)}

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

Table 5
Trained belief rules for Example 2.

Rule no.	Rule weight	x	$g(x)$	Belief degrees on $\{D_1, D_2, D_3, D_4, D_5\}$
1	0.9998	−5	−0.0035	(0.1411, 0.8037, 0.0000, 0.0320, 0.0233)
2	0.3860	−2.0012	0.4986	(0.1067, 0.0414, 0.7016, 0.0486, 0.1017)
3	0.5463	0.0532	0.0365	(0.0396, 0.8852, 0.0377, 0.0375, 0.0000)
4	0.3798	2.0314	0.9897	(0.0653, 0.0325, 0.0109, 0.6402, 0.2511)
5	1.0000	5	−0.0119	(0.1280, 0.8112, 0.0391, 0.0000, 0.0217)

As shown in Fig. 3, it looks that the inference output generated by the trained belief rules can closely replicate the nonlinear relationship of the multi-extremal function. In fact, the MSE between the actual output and the inference output is only 6.32×10^{-5}. The attribute weight keeps constant at 1, and the locally decomposed regions are slightly updated to $-5 \leqslant x \leqslant -2.0012$, $-2.0012 \leqslant x \leqslant 0.0532$, $0.0532 \leqslant x \leqslant 2.0314$ and $2.0314 \leqslant x \leqslant 5$. Table 5 lists the trained belief rules in detail.

The analytical and numerical analyses of this section reveal the inference mechanisms of BRB systems, leading to a conclusion that a BRB system combines a multi-model decomposition modelling structure with a distributed approximation process, and can approximate complex nonlinear causal relationships between antecedent inputs and output.

4. Universal approximation of BRB systems

In this section, we further prove that BRB systems are universal approximators. In other words, BRB systems can approximate any real continuous function on a compact set with arbitrary accuracy.

To demonstrate that BRB systems are capable of providing close approximation to any continuous systems, we first give a brief introduction to a new universal approximation theorem for BRB systems. In the following, let $F(x)$ be the set of analytical models produced by the unified inference structure of BRB systems given in Eqs. (3)–(7), and $U \subseteq R^m$ be a compact set of input variables denoted by $[A_{1,1}, A_{1J_1}] \times \cdots \times [A_{M,1}, A_{1J_M}]$.

Theorem 1. Universal approximation theorem*For any given real continuous function g(x) on a compact domain $U \subseteq R^m$ and arbitrary $\varepsilon > 0$, there exists a BRB system $f(x) \in F(x)$ with $F(x)$ being the set of all BRB systems, such that*

$$\sup_{x \in U} |g(x) - f(x)| < \varepsilon$$

The proof of the universal approximation theorem using Stone–Weierstrass theorem will be presented later. First of all, we provide an intuitive discussion on the approximation capability of BRB systems. Since the continuity of $g(x)$ on the compact domain U guarantees its uniform continuity, there exists a fixed distance δ or a set of δ_i, $i = 1, \ldots, M$ such that, for all x and x' in U, $|g(x) - g(x')| < \varepsilon$ if $|x - x'| < \delta$ or $|x_i - x_i'| < \delta_i, \forall i = 1, \ldots, M$. Thus we can construct a BRB system in which the referential values of antecedent input x_i can be defined as $A_{i,j} = A_{i,1} + (j-1) \times (A_{i,J_i} - A_{i,1})/J_i$ with $J_i = \lfloor (A_{i,J_i} - A_{i,1})/\delta_i \rfloor$. It means that we can make the approximation error arbitrarily small by choosing a large number of referential values, i.e. increasing J_i. From the discussion in Section 3.2, a small number of dedicated belief rules can actually have the power of approximating complicated nonlinear systems.

Strictly, the following Stone–Weierstrass theorem [4] is used to prove the universal approximation property of BRB systems.

Theorem 2. (Stone–Weierstrass theorem). *Let $F(x)$ be a set of real continuous functions on a compact domain $U \subseteq R^m$. If (1) F is an algebra, i.e., the set F is closed under addition, multiplication, and scalar multiplication, (2) vanishes at no point of$U$, i.e., for each $x \in U$ there exists $f \in F$ such that $f(x) \neq 0$, and (3) F separates points on U, i.e., for every $x, x' \in U$, $x \neq x'$, there exists $f \in F$ such that $f(x) \neq f(x')$, then the uniform closure of F consists of all real continuous functions on U.*

First of all, we study the necessary preconditions of the Stone–Weierstrass theorem. To simplify expressions, an intermediate variable is defined as follows:

$$\chi_n(x) = \prod_{k=1}^{K} (w_k(x)\beta_{n,k} + 1 - w_k(x)) - \prod_{k=1}^{K} (1 - w_k(x)) \tag{14}$$

The numerical output of a BRB system can then be represented as

$$f(x) = \frac{\sum_{n=1}^{N} u(D_n)\chi_n(x)}{\sum_{n=1}^{N} \chi_n(x)} \tag{15}$$

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

Lemma 5. *The set of all BRB systems F(x) is an algebra. In other words, F(x) is closed under addition, multiplication, and scalar multiplication.*

Proof. Let $f_1, f_2 \in F$, so that we can calculate their sum and product as follows:

$$f_1(x)+f_2(x)=\frac{\sum_{n1=1}^{N1}u^1(D_{n1})\chi_{n1}^1(x)}{\sum_{n1=1}^{N1}\chi_{n1}^1(x)}+\frac{\sum_{n2=1}^{N2}u^2(D_{n2})\chi_{n2}^2(x)}{\sum_{n2=1}^{N2}\chi_{n2}^2(x)}=\frac{\sum_{n1=1}^{N1}\sum_{n2=1}^{N2}\left[u^1(D_{n1})+u^2(D_{n2})\right]\chi_{n1}^1(x)\chi_{n2}^2(x)}{\sum_{n1=1}^{N1}\sum_{n2=1}^{N2}\chi_{n1}^1(x)\chi_{n2}^2(x)} \tag{16}$$

$$f_1(x)f_2(x)=\frac{\sum_{n1=1}^{N1}\sum_{n2=1}^{N2}u^1(D_{n1})u^2(D_{n2})\chi_{n1}^1(x)\chi_{n2}^2(x)}{\sum_{n1=1}^{N1}\sum_{n2=1}^{N2}\chi_{n1}^1(x)\chi_{n2}^2(x)} \tag{17}$$

It is clear from Eq. (14) that $\chi_n(x)$ can be represented as a polynomial form of the activation weight $w_k(x)$. Therefore, the product $\chi_{n1}^1(x)\chi_{n2}^2(x)$ can also be presented as a polynomial form of the activation weight product $w_{k1}^1(x)w_{k2}^2(x)$. Since the activation weights $w_{k1}^1(x)$ and $w_{k2}^2(x)$ are calculated by the normalised multiplicative matching function in Eq. (4), their product $w_{k1}^1(x)w_{k2}^2(x)$ also have the same form as Eq. (4). In fact, if the attribute weights of all antecedent attributes are set to be 1, Eq. (14) can be represented by a polynomial form of the input x. Allowing attribute weights to be adjustable further enhances the approximation capability of BRB systems. As a conclusion, Eqs. (16) and (17) can be transformed to the same form as Eq. (15), which proves that $f_1 + f_2 \in F$ and $f_1 f_2 \in F$.

In addition, for arbitrary $c \in R$,

$$cf(x)=\frac{\sum_{n=1}^{N}c\cdot u(D_n)\chi_n(x)}{\sum_{n=1}^{N}\chi_n(x)} \tag{18}$$

which is in the same form as Eq. (15), which proves that $cf \in F$.

Therefore, we can conclude that $F(x)$ is closed under addition, multiplication, and scalar multiplication.

Lemma 6. *.For each $x \in U$, there exists a BRB system $f \in F$ such that $f(x) \neq 0$.*

Proof. By analysing Eqs. (5) and (7), we can simply construct a BRB system in which $u(D_N) > 0$, $\beta_{N,k} = 1$, $k = 1, \ldots, K$, and $\beta_{n,k} = 0$, $\forall n \neq N$. The output of the BRB system will be equal to $u(D_N)$ for each $x \in U$.

Lemma 7. *For every $x,x' \in U$, $x \neq x'$, there exists a BRB system $f \in F$ such that $f(x) \neq f(x')$.*

Proof. This can be proved by constructing a required BRB system. For simplicity and without loss of generality, we construct a BRB system with two rules to model a single input system, represented as follows:

$$R_k : \{(D_n, \beta_{n,k}), n = 1, \ldots, N\} \quad \text{with } \beta_{1,k} = 1$$
$$R_l : \{(D_n, \beta_{n,l}), n = 1, \ldots, N\} \quad \text{with } \beta_{N,l} = 1$$

Suppose the antecedent of R_k is the lower bound A_1 of the input variable x, and that of R_k is the upper bound A_2. Without loss of generality, we assume that $x < x'$ for every $x, x' \in U$, $x \neq x'$, and $u(D_N) > u(D_1)$. According to Eqs. (3) and (4), we can get $w_k(x) > w_k(x')$. It is also straightforward to prove that $\frac{df(x)}{dw_k(x)} < 0$ on the basis of Eq. (5). Hence, we conclude that $f(x) < f(x')$ for this designed BRB system, or $f(x) \neq f(x')$.

By using the Stone–Weierstrass theorem together with Lemmas 5–7, we prove that a BRB system possesses the universal approximation capability.

Remark 3. The universal approximation theorem shows that a BRB system can approximate any real continuous function over a compact subset of R^m if it decomposes input space using a sufficiently large number of referential values for antecedent attributes. For example, for complex system identification problems, this means that a BRB system can uniformly approximate any given continuous output trajectory $y(t)$ of any nonlinear dynamic system over any compact time-interval $t \in [t_0, T]$ with arbitrarily high accuracy. In addition, the proof also shows that the change of parameters may lead to significant changes in the approximation capability of a BRB system. However, this universal approximation theorem just guarantees the existence of a BRB system which is capable of approximating a given continuous system with arbitrary accuracy. In designing a practical BRB system, however, we still have to exploit experts' knowledge and historical data, and make tradeoffs between approximation accuracy and complexity for training and inference.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

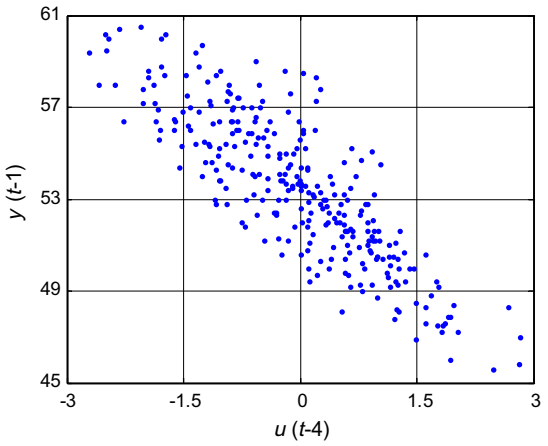

Fig. 4. Training data in decomposed input space.

5. A numerical simulation study

In this section, a numerical simulation study is conducted on the system identification of Box–Jenkins gas furnace [2]. It is a well-known benchmark system identification problem, and many researchers have used the data set to demonstrate their nonlinear identification methods [7,14,15,19]. The original data set was recorded from a combustion process of a methane-air mixture, and consists of 296 data pairs $\{(u(t), y(t));\ t = 1, \ldots, 296\}$, where $u(t)$ is the input gas flow rate, and $y(t)$ is the observed output CO_2 concentration at the sampling time t [2]. In this paper, the following series parallel model is employed to formulate the system dynamics [14].

$$y(t) = f(y(t-1), u(t-4)) \tag{19}$$

Thus there will be 292 input–output data pairs. To construct a BRB system for this system identification problem, we use five linguistic terms to define $u(t)$, which are negative large (NL), negative small (NS), zero (Z), positive small (PS), and positive large (PL). These linguistic terms are associated with numerical referential values, which are given by

$$A_1^k \in \{\mathrm{NL}, \mathrm{NS}, \mathrm{Z}, \mathrm{PS}, \mathrm{PL}\} = \{-3, -1.5, 0, 1.5, 3\}$$

Similarly, we also use 5 linguistic terms to define $y(t)$, which are Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH) and are associated with the following referential values.

$$A_2^k \in \{\mathrm{VL}, \mathrm{L}, \mathrm{M}, \mathrm{H}, \mathrm{VH}\} = \{45, 49, 53, 57, 61\}$$

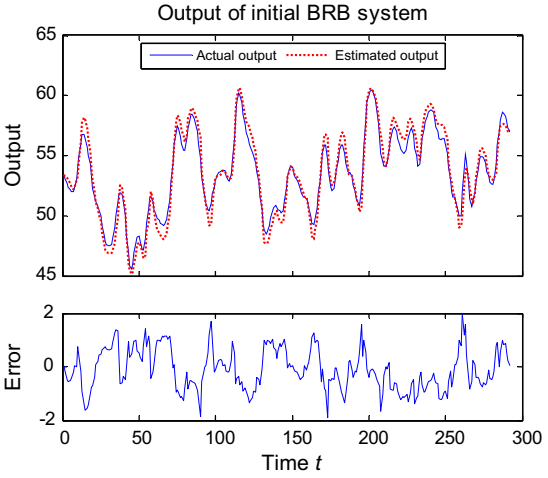

Fig. 5. Comparison of actual output and initial BRB output.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

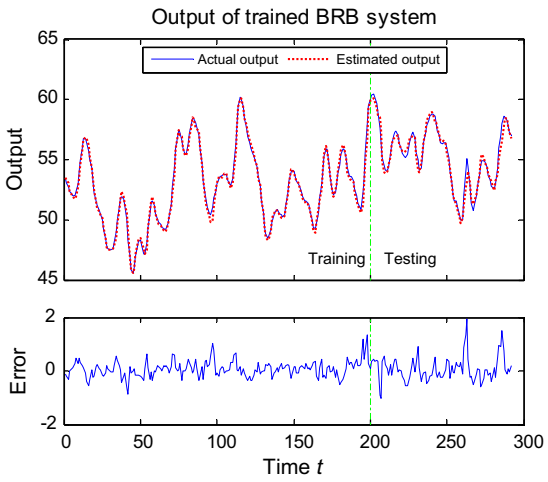

Fig. 6. Estimated output of the trained BRB system.

Given the above definition of the referential values, there are 25 referential combinations of the two antecedent inputs $y(t-1)$ and $u(t-4)$, leading to 25 belief rules in total in the belief rule base. As discussed in Section 3.1, the referential values decompose the input space into a set of 16 local regions. All the 292 input–output data pairs fall into the 16 different local regions as shown in Fig. 4.

By examining these historical data and using experts' knowledge that the output CO_2 concentration decreases with a time delay if the input gas flow rate increases, the initial belief rules can be constructed to model the nonlinear causal relationships between the antecedent inputs $(u(t-4), y(t-1))$ and the output $y(t)$. Table A-I lists the initial belief rules. The output $y(t)$ is characterised by a belief distribution $\{D_1, D_2, D_3, D_4, D_5\} = \{45, 49, 53, 57, 61\}$ in each rule. As discussed in Section 2, we can use the constructed belief rules to infer the estimated output for each sampling input data $(u(t-4), y(t-1))$. Fig. 5 shows the comparison of the actual output and the estimated output inferred by the initial BRB system.

It can be said from Fig. 5 that the estimated output using the initial BRB system without training is not too bad but the approximation error is relatively large in some areas. This is because the parameters in the initial rule base are not very accurate. It is therefore necessary to train the parameters of the initial BRB system. In our simulation, the first portion of the time-series dataset consisting of 200 data pairs is used to train the BRB system, and the remaining part of the dataset is used to test the training effects for validation purpose. The MSE defined in Eq. (8) is used to be the optimal learning objective, and constraints are discussed in Refs. [5,32]. Note that in Fig. 5 the MSE between the actual output and the initial BRB output is 0.5942. By using the nonlinear optimisation solver *fmincon* in the Optimisation Toolbox of Matlab, we can further obtain the trained belief rule base with the rule weights as listed in Table A-II. The trained attribute weights of $u(t-4)$ and $y(t-1)$ are equal to 0.85 and 0.77. The referential values are updated slightly, given by {-3, −1.5, 0, 1.5, 3} and {45, 49.17, 53.15, 57.07, 61} respectively. Fig. 6 shows the estimated output of the trained BRB system for both the training and testing datasets.

As shown in Fig. 6, the trained BRB system has much better inference performance than the initial BRB system, and the relative error has been reduced significantly with the optimal MSE of 0.0864. We can see from Table A-II that the belief degrees and rule weights are also updated for improving the approximation accuracy of the BRB system. Table 6 reports the approximation results using the BRB systems with different numbers of referential values. Note that for each BRB system the initial referential values of the two antecedent inputs are uniformly sampled from the intervals [-3, 3] and [45, 61] respectively.

As one can see in Table 6, the MSE of the training data set is monotonically decreasing with the increase of the number of referential values. This is consistent with the theoretical analysis in Section 4. The MSE of the testing data set depends on the generalisation capability of each trained BRB system. Even for the BRB system with the number of 2×2 referential values, the approximation errors of both the training data set and the testing data set are as small as or smaller than those reported in [7,15,19]. It should be pointed out that the experimental settings, such as the number of parameters and the number of

Table 6
Approximation results using the BRB systems with different numbers of referential values.

Approx.	No. of referential values	MSE (training data set)	MSE (testing data set)
1	2×2	0.1190	0.2403
2	3×3	0.1042	0.2262
3	4×4	0.1032	0.2284
4	5×5	0.0864	0.2131

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

Table A-I
Initial belief rule base ($\{D_1, D_2, D_3, D_4, D_5\} = \{45, 49, 53, 57, 61\}$).

Rule no.	Rule weight	$u(t-4)$ AND $y(t-1)$	Consequents $\{D_1, D_2, D_3, D_4, D_5\}$
1	1	NL ∧ VL	$\{(D_1, 0.1), (D_2, 0.9), (D_3, 0), (D_4, 0), (D_5, 0)\}$
2	1	NL ∧ L	$\{(D_1, 0), (D_2, 0.5), (D_3, 0.5), (D_4, 0), (D_5, 0)\}$
3	1	NL ∧ M	$\{(D_1, 0), (D_2, 0), (D_3, 0.5), (D_4, 0.5), (D_5, 0)\}$
4	1	NL ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.3), (D_5, 0.7)\}$
5	1	NL ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 1)\}$
6	1	NS ∧ VL	$\{(D_1, 0.25), (D_2, 0.75), (D_3, 0), (D_4, 0), (D_5, 0)\}$
7	1	NS ∧ L	$\{(D_1, 0), (D_2, 0.75), (D_3, 0.25), (D_4, 0), (D_5, 0)\}$
8	1	NS ∧ M	$\{(D_1, 0), (D_2, 0), (D_3, 0.5), (D_4, 0.5), (D_5, 0)\}$
9	1	NS ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.7), (D_5, 0.3)\}$
10	1	NS ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.25), (D_5, 0.75)\}$
11	1	Z ∧ VL	$\{(D_1, 0.5), (D_2, 0.5), (D_3, 0), (D_4, 0), (D_5, 0)\}$
12	1	Z ∧ L	$\{(D_1, 0), (D_2, 0.9), (D_3, 0.1), (D_4, 0), (D_5, 0)\}$
13	1	Z ∧ M	$\{(D_1, 0), (D_2, 0), (D_3, 0.95), (D_4, 0.05), (D_5, 0)\}$
14	1	Z ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4,1), (D_5, 0)\}$
15	1	Z ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.5), (D_5, 0.5)\}$
16	1	PS ∧ VL	$\{(D_1, 0.75), (D_2, 0.25), (D_3, 0), (D_4, 0), (D_5, 0)\}$
17	1	PS ∧ L	$\{(D_1, 0.3), (D_2, 0.7), (D_3, 0), (D_4, 0), (D_5, 0)\}$
18	1	PS ∧ M	$\{(D_1, 0), (D_2, 0.35), (D_3, 0.65), (D_4, 0), (D_5, 0)\}$
19	1	PS ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0.5), (D_4, 0.5), (D_5, 0)\}$
20	1	PS ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.75), (D_5, 0.25)\}$
21	1	PL ∧ VL	$\{(D_1,1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
22	1	PL ∧ L	$\{(D_1, 0.75), (D_2, 0.25), (D_3, 0), (D_4, 0), (D_5, 0)\}$
23	1	PL ∧ M	$\{(D_1, 0), (D_2, 0.75), (D_3, 0.25), (D_4, 0), (D_5, 0)\}$
24	1	PL ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0.75), (D_4, 0.25), (D_5, 0)\}$
25	1	PL ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4,1), (D_5, 0)\}$

Table A-II
Trained belief rule base ($\{D_1, D_2, D_3, D_4, D_5\} = \{45, 49.17, 53.15, 57.07, 61\}$).

Rule no.	Rule weight	$u(t-4)$ AND $y(t-1)$	Consequents $\{D_1, D_2, D_3, D_4, D_5\}$
1	1	NL ∧ VL	$\{(D_1, 0.1), (D_2, 0.9), (D_3, 0), (D_4, 0), (D_5, 0)\}$
2	1	NL ∧ L	$\{(D_1, 0), (D_2, 0.5), (D_3, 0.5), (D_4, 0), (D_5, 0)\}$
3	0.84	NL ∧ M	$\{(D_1, 0.26), (D_2, 0.12), (D_3, 0.27), (D_4, 0.11), (D_5, 0.24)\}$
4	0.99	NL ∧ H	$\{(D_1, 0), (D_2, 0.13), (D_3, 0), (D_4, 0.1), (D_5, 0.77)\}$
5	1.00	NL ∧ VH	$\{(D_1, 0.04), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.96)\}$
6	1.00	NS ∧ VL	$\{(D_1, 0.25), (D_2, 0.75), (D_3, 0), (D_4, 0), (D_5, 0)\}$
7	1.00	NS ∧ L	$\{(D_1, 0.24), (D_2, 0.18), (D_3, 0.35), (D_4, 0.13), (D_5, 0.1)\}$
8	0.31	NS ∧ M	$\{(D_1, 0.08), (D_2, 0.2), (D_3, 0), (D_4, 0.72), (D_5, 0)\}$
9	0.51	NS ∧ H	$\{(D_1, 0.04), (D_2, 0), (D_3, 0.01), (D_4, 0.89), (D_5, 0.06)\}$
10	0.62	NS ∧ VH	$\{(D_1, 0.08), (D_2, 0.01), (D_3, 0.07), (D_4, 0.01), (D_5, 0.83)\}$
11	0.84	Z ∧ VL	$\{(D_1, 0.55), (D_2, 0.32), (D_3, 0.01), (D_4, 0.07), (D_5, 0.05)\}$
12	0.57	Z ∧ L	$\{(D_1, 0), (D_2, 0.77), (D_3, 0.03), (D_4, 0.2), (D_5, 0)\}$
13	0.39	Z ∧ M	$\{(D_1, 0.31), (D_2, 0), (D_3, 0), (D_4, 0.68), (D_5, 0.01)\}$
14	0.27	Z ∧ H	$\{(D_1, 0.28), (D_2, 0.09), (D_3, 0), (D_4, 0), (D_5, 0.63)\}$
15	1.00	Z ∧ VH	$\{(D_1, 0.06), (D_2, 0.03), (D_3, 0), (D_4, 0.25), (D_5, 0.66)\}$
16	0.84	PS ∧ VL	$\{(D_1, 0.65), (D_2, 0.15), (D_3, 0), (D_4, 0.2), (D_5, 0)\}$
17	0.61	PS ∧ L	$\{(D_1, 0.58), (D_2, 0.16), (D_3, 0), (D_4, 0.26), (D_5, 0)\}$
18	0.66	PS ∧ M	$\{(D_1, 0.19), (D_2, 0), (D_3, 0.81), (D_4, 0), (D_5, 0)\}$
19	0.40	PS ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.46), (D_5, 0.54)\}$
20	1.00	PS ∧ VH	$\{(D_1, 0.49), (D_2, 0.17), (D_3, 0.02), (D_4, 0.3), (D_5, 0.02)\}$
21	1.00	PL ∧ VL	$\{(D_1, 0.95), (D_2, 0.03), (D_3, 0), (D_4, 0.02), (D_5, 0)\}$
22	0.40	PL ∧ L	$\{(D_1, 0.77), (D_2, 0.01), (D_3, 0), (D_4, 0.04), (D_5, 0.18)\}$
23	0.99	PL ∧ M	$\{(D_1, 0.22), (D_2, 0.72), (D_3, 0.06), (D_4, 0), (D_5, 0)\}$
24	1	PL ∧ H	$\{(D_1, 0), (D_2, 0), (D_3, 0.75), (D_4, 0.25), (D_5, 0)\}$
25	1	PL ∧ VH	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4,1), (D_5, 0)\}$

data samples used for testing, are different in each paper. As the BRB system and fuzzy neural models have different structures, it is difficult to conduct a numerical comparison under the same experiment setting, but it can be an advantage to use more training samples.

6. Conclusion and discussion

In this paper, theoretical, analytical and numerical analyses were conducted about the inference and approximation properties of BRB systems. Firstly of all, the unified multi-model decomposition structure of BRB systems was analysed. In a spe-

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

cific BRB system, its referential values for antecedent attributes decompose the input space into different local regions. In a local region, its inference output is generated by the distributed approximation process. These underlying inference mechanisms enable BRB systems to have superior approximation capability in practical applications. Furthermore, a constructive proof was given that a BRB system is capable of approximating any continuous function on a compact set with arbitrary accuracy. This universal approximation property ensures that a BRB system with adaptively learning parameters can be applied to approximate complex nonlinear systems.

In comparison with some state-of-the-art methodologies, such as statistical methods, fuzzy rule-based systems, ANN and support vector machine (SVM), which can also be used to model complex nonlinear systems, the BRB methodology has the following features [31,32,38]:

(1) BRB models are transparent and their input–output relationships characterised by belief rules are interpretable, while ANN and SVM models are black boxes in nature and their input–output relationships are difficult to be interpreted;
(2) Expert knowledge can be incorporated into BRB models as belief rules or constraints explicitly. In addition, the initial values of the training parameters in BRB models can be set by experts intuitively whenever possible, leading to an expert-data driven BRB system. However, statistical methods, ANN and SVM models have no such capability. Fuzzy rule-based systems can also be used to model expert knowledge, but it is not always clear how to construct appropriate membership functions for type-1 fuzzy systems and how to interpret various fuzzy aggregation operators for both type-1 and type-2 fuzzy systems [39,40]. But it worth noting that interval type-2 fuzzy sets are useful to solve the problem when defining the membership functions.
(3) In BRB models, the output can be either a single value or a belief distribution, while statistical methods, ANN and SVM models can only use average values as outputs and have no way to model a belief distribution;
(4) BRB models can be used to handle uncertainty such as incomplete information and also set targets for antecedent inputs and intermediate variables, while statistical methods, ANN and SVM models do not have such capability;
(5) BRB models can identify what belief rules cannot be learned and what samples cannot be predicted with confidence if any, but other existing models cannot provide such information and may provide false predictions that cannot be learned from the training samples.

However, BRB system is not developed to compete against these methodologies in situations where a complete range of high-quality data samples is available. It has greater potential to be applied to more general approximation situations with various uncertainties, and has shown the capability of capturing complicated nonlinear causal relationships between antecedent attributes and consequents and also incorporating expert's subjective knowledge in several real-world applications [5,27,36,38].

Finally, in the paper a numerical simulation study on the well-known nonlinear system identification problem of Box–Jenkins gas furnace was conducted, which illustrates that the trained BRB system has superior knowledge representation and inference capability. This investigation serves to promote the further development of knowledge-based systems for solving various identification and decision problems in complex nonlinear systems. However, further study is needed in the following aspects.

(1) Although BRB systems are proved to be general approximators, the universal approximation theorem does not show how to construct a BRB system to approximate a specific complex nonlinear system and how many belief rules would be needed to achieve given approximation accuracy. The optimal design of BRB systems needs to be studied.
(2) In this paper the inference and approximation properties of single-layer BRB systems were analysed. In single-layer BRB systems, the total number of rules increases exponentially with the increase of the number of input variables. In practice, hierarchical BRB systems have been used to deal with the "curse of dimensionality". The inference mechanisms and approximation capability of hierarchical BRB systems need to be studied.
(3) With the universal approximation capability, any lack of success in the applications of BRB systems can be attributed to inadequate learning, an insufficient number of rules, or the lack of a deterministic relationship between antecedent inputs and output. Therefore, the ways of properly constructing belief rule bases and effectively learning design parameters should be considered carefully in real-world applications.

Acknowledgements

This work forms part of the project supported by the UK Engineering and Physical Science Research Council under Grant No. EP/F024606/1 and the Natural Science Foundation of China under Grant No. 60736026. We would like to thank the editor and the three anonymous referees for the comments and suggestions which have substantially improved the quality of the paper.

Appendix A. Belief rule bases of Box–Jenkins gas furnace

Tables A-I and A-II.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022

References

[1] J.C. Augusto, J. Liu, P.J. McCullagh, H. Wang, J.B. Yang, Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a smart home, International Journal of Computational Intelligence Systems 1 (4) (2008) 361–378.
[2] G.E.P. Box, G.M. Jenkins, G.C. Reinse, Time Series Analysis: Forecasting and Control, Prentice Hall, NJ, 1994.
[3] J.L. Castro, Fuzzy logic controllers are universal approximators, IEEE Transactions on Systems, Man, and Cybernetics 25 (4) (1995) 629–635.
[4] E.W. Cheney, Introduction to Approximation Theory, AMS Chelsea Publishing, 1998.
[5] Y.W. Chen, J.B. Yang, D.L. Xu, Z.J. Zhou, D.W. Tang, Inference analysis and adaptive training for belief rule based systems, Expert Systems with Applications 36 (10) (2011) 12845–12860.
[6] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society, Series B 30 (2) (1968) 205–247.
[7] X. Deng, X. Wang, Incremental learning of dynamic fuzzy neural networks for accurate system modeling, Fuzzy Sets and Systems 160 (2009) 972–987.
[8] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence 172 (2008) 234–264.
[9] P. Giovanni, I. Lurdes, Decision Theory: Principles and Approaches, John Wiley & Sons, 2009.
[10] A. Gegov, Distributed Fuzzy Control of Multivariable Systems, Kluwer Academic Publishers, The Netherlands, Dordrecht, 1996.
[11] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5) (1989) 359–366.
[12] F. Hayes-Roth, Rule-based systems, Communications of the ACM 28 (9) (1985) 921–932.
[13] V. Kecman, Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models, MIT Press, Cambridge, MA, 2001.
[14] J. Kim, N. Kasabov, HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems, Neural Networks 12 (1999) 1301–1319.
[15] D. Kukolj, E. Levi, Identification of complex systems based on neural and Takagi–Sugeno fuzzy model, IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics 34 (1) (2004) 272–282.
[16] J. Lawry, A framework for linguistic modelling, Artificial Intelligence 155 (2004) 1–39.
[17] A. Ligęza, Logical Foundations for Rule-Based Systems, Springer-Verlag, Berlin, Heidelberg, 2006.
[18] J. Liu, J.B. Yang, D. Ruan, L. Martinez, J. Wang, Self-tuning of fuzzy belief rule bases for engineering system safety analysis, Annals of Operations Research 163 (1) (2008) 143–168.
[19] B. Rezaee, M.H. Fazel Zarandi, Data-driven fuzzy modeling for Takagi–Sugeno–Kang fuzzy system, Information Sciences 180 (2010) 241–255.
[20] G. Shafer, Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[21] M.H. Stone, The generalized Weierstrass approximation theorem, Mathematics Magazine 21 (1948) 167–183.
[22] R. Sun, Robust reasoning: integrating rule-based and similarity-based reasoning, Artificial Intelligence 75 (1995) 241–295.
[23] D. Tikk, L.T. Kóczy, T.D. Gedeon, A survey on universal approximation and its limits in soft computing techniques, International Journal of Approximate Reasoning 33 (2003) 185–202.
[24] L.X. Wang, J.M. Mendel, Fuzzy basis functions, universal approximation, and orthogonal least-squares learning, IEEE Transactions on Neural Networks 3 (5) (1992) 807–814.
[25] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, European Journal of Operational Research 174 (3) (2006) 1885–1913.
[26] D.L. Xu, J.B. Yang, Y.M. Wang, The ER approach for multi-attribute decision analysis under interval uncertainties, European Journal of Operational Research 174 (3) (2006) 1914–1943.
[27] D.L. Xu, J. Liu, J.B. Yang, G.P. Liu, J. Wang, I. Jenkinson, J. Ren, Inference and learning methodology of belief-rule-based expert system for pipeline leak detection, Expert Systems with Applications 32 (1) (2007) 103–113.
[28] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple attribute decision making with uncertainty, IEEE Transactions on Systems, Man, and Cybernetics 24 (1) (1994) 1–18.
[29] J.B. Yang, Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty, European Journal of Operational Research 131 (1) (2001) 31–61.
[30] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 32 (3) (2002) 289–304.
[31] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, A belief rule-base inference methodology using the evidential reasoning approach - RIMER, IEEE Transactions on Systems, Man, and Cybernetics – Part A 36 (2) (2006) 266–285.
[32] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimization models for training belief rule based systems, IEEE Transactions on Systems, Man, and Cybernetics – Part A 37 (4) (2007) 569–585.
[33] K. Yoon, C.L. Hwang, Multiple Attribute Decision Making: An Introduction, Sage Publications, Thousand Oaks, CA, 1995.
[34] X.J. Zeng, M.G. Singh, Approximation accuracy analysis of fuzzy systems as function approximators, IEEE Transactions on Fuzzy Systems 4 (1) (1996) 44–53.
[35] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief-rule-based system for pipeline leak detection under expert intervention, Expert Systems with Applications 36 (2009) 7700–7709.
[36] Z.J. Zhou, C.H. Hu, D.L. Xu, J.B. Yang, D.H. Zhou, New model for system behaviour prediction based on belief-rule-based systems, Information Sciences 180 (2010) 4834–4864.
[37] V.A. Zorich, Mathematical Analysis, Springer-Verlag, Berlin, Heidelberg, 2004.
[38] J.B. Yang, Y.M. Wang, D.L. Xu, K.S. Chin, L. Chatton, Belief rule-based methodology for mapping consumer preferences and setting product targets, Expert Systems with Applications 39 (2012) 4749–4759.
[39] J.M. Mendel, Type-2 fuzzy sets and systems: an overview, IEEE Computational Intelligence Magazine (2007) 20–29.
[40] J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning, Information Sciences 180 (19) (2010) 3674–3685.

Please cite this article in press as: Y.-W. Chen et al., On the inference and approximation properties of belief rule based systems, Inform. Sci. (2013), http://dx.doi.org/10.1016/j.ins.2013.01.022