IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Belief rule base inference method based on gradient descent with momentum

FIRST A. AUTHOR[1], (Fellow, IEEE), SECOND B. AUTHOR[2], AND THIRD C. AUTHOR, JR.[3], (Member, IEEE)

[1]National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov)
[2]Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu)
[3]Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA

Corresponding author: First A. Author (e-mail: author@ boulder.nist.gov).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT The belief-rule-base(BRB) inference methodology using evidential reasonging(ER) approach is widely used in different fields, such as fault diagnosis, system identification and decision analysis. In this paper, we propose a new belief rule structure and its training method, aiming to solve zero activation during the inference process and improve inference accuray. We first used the Gaussian function to calculate the similarity of each attribute instead of the original method. Then we introduce corresponding attribute weight for each rule and cancel the rule weight parameter at the same time. Finally, we use the stochastic gradient descent method for parameters training based on the new rule structure. Experiments on several public classification datasets are conducted to validate the proposed approach compared with some recent existing works. The experimental results show that the proposed approach have a better performence in accuray and time consumption.

INDEX TERMS belief rule base, structure optimization, stochastic gradient descent, momentum optimization.

I. INTRODUCTION

The belief rule-based inference methodology using evidential reasonging approach(RIMER) proposed by Yang [?] baesd on traditional IF-THEN rules [?], Dempster-Shafer theory of evidence [?], [?], decision theory [?] and fuzzy set theory [?]. By introducing a belief distribution structure in the rules, this methodology can effectively handle incomplete and uncertain information involved in the datasets and widely used in various problem in different fields such as oil pipeline leak detection [?], military capability estimation [?], consumer behavior prediction [?] and so on.

In the inference process of the BRB system, the attribute weight, rule weight, belief distribution and other parameters directly affect the final accuray. Yang [?] proposed optimization models for training BRB system using fmincon solver in Matlab, Chang [?], [?] proposed an algorithm for training parameters in BRB system based on gradient and dichotomy methods, Wu [?] used the accelerating of gradient algorithm to improve the convergence accuray and convergence speed. There are also a series of intelligent algorithms such as the particle swarm algorithm proposed by Su [?] and the differential evolution algorithm proposed by Wang [?] have excellent training effects on the BRB system. Liu [?] introduces the belief distribution structure into the antecedent attributes and uses training data to build an extended belief rule base(EBRB) system, which simplifies the construction of the rule base and improves the inference speed.

At present, the parameter optimization model of the BRB system is mostly based on various intelligent algorithms. Their process is complicated and there are many intermediate training parameters. When the traditional gradient method is used to train the parameters of the BRB system, the step size is restricted by a variety of constraints, and other methods are needed to find the optimal step size. The EBRB system does not introduce a parameter training process, which makes the system have higher requirements for the representativeness of the training data selected to build the rule base. In the case of a large number of rules, it is necessary to perform rule reduction or use the data structure to optimize the storage and activation process of the rules. Because the traditional

BRB system includes the rule attribute reference level setting, its potential zero activation problem may cause the inference system to fail.

In response to the above problems, we have proposed a series of optimization modifications to the system structure and reasoning process, including:

1) We propose a new antecedent structure that does not need to set the attribute reference level, and proposed a Gaussian function-based rule weight activation method for the new rule antecedent structure, which can effectively avoid the zero activation problem and has the feature of generating rules from the training data like EBRB.

2) We change the method of setting the weight of the global same antecedent attribute in the traditional BRB system, and set the corresponding rule attribute weight parameter for each rule, so that each rule has a finer activation granularity. On this basis, the rule weight and its related normalization process are cancelled, which simplifies the evidential reasoning process.

3) We further introduce the linear rectification function and the normalized exponential function to preprocess the restricted parameters to avoid the problem of parameter failure during the training process.

The remainder of this paper is organized as follows: Section II introduces the traditional BRB system and our further improvements for common problems in the system. In Section III, we give the preprocessing method of the training model and prove that the gradient descent method can be effectively applied to the newly proposed BRB system. In Section IV, we compare the effects of different gradient descent parameters on training speed and inference accuracy. Experiments on a series of public data sets prove that the newly proposed BRB model and its training method have excellent performance. Section V concludes this paper.

II. BRB SYSTEM WITH NEW ATTRIBUTE STRUCTURE AND RULE ACTIVATION WEIGHT CALCULATION METHOD

The BRB system proposed by Yang mainly refers to the rule activation and evidence reasoning method on the belief rule base. This part will briefly introduce the related concepts of the BRB system and propose solutions to the common defects of the traditional BRB system.

A. REPRESENTATION OF BELIEF RULE BASE

On the basis of the traditional production rules, Yang [?] proposed the expression form of the belief rules by introducing the belief distribution structure, the rule antecedent attribute parameter and the rule weight parameter. The specific expression is as follows:

$$R_k : if\{X_1 is A_1^k \wedge \cdots \wedge X_{T_k} is A_{T_k}^k\}$$

$$then\{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\}, \sum_{i=1}^{N} \beta_i^k \leq 1$$

The equal sign is obtained when the rule information is complete. Each rule has its rule weight θ_k, antecedent attribute weight $\delta_1, \delta_2, \cdots, \delta_{T_k}$. A_i^k represents the candidate reference value selected by the rule on the i-th attribute and β_i^k represents the belief degree of the rule in the i-th result attribute. On this basis, the extended belief rule base system introduces a belief distribution structure to the antecedent attributes, and its rule form is expressed as follows:

$$R_k : if\{[(A_{11}^k, \alpha_{11}^k), \cdots, (A_{1J_1}^k, \alpha_{1J_1}^k)] \wedge$$
$$\cdots \wedge [(A_{T_k 1}^k, \alpha_{T_k 1}^k), \cdots, (A_{T_k J_{T_k}}^k, \alpha_{T_k J_{T_k}}^k)]\}$$

$$then\{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\}, \sum_{i=1}^{N} \beta_i^k \leq 1$$

The extended belief rule base constructed using the original data transforms the input data into the rule antecedent attributes in the form of belief distribution. For the input data:

$$X^k = (x_1^k, \cdots, x_T^k)$$

Convert the ith attribute parameter to construct the ith antecedent attribute of the corresponding rule with a belief distribution form:

$$\alpha_{ij}^k = \frac{\gamma_{i(j+1)-x_i^k}}{\gamma_{i(j+1)-\gamma_{ij}}}, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)}$$

$$\alpha_{i(j+1)}^k = 1 - \alpha_{ij}^k, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)}$$

$$\alpha_{it}^k = 0, t = 1, \cdots, (j-1), (j+2), \cdots, J_i$$

According to the same conversion method, the values of original data on other attributes can be converted into the corresponding berlief distribution form. We can also obtain the belief distribution form of the rule result attribute according to this method.

B. EVIDENCE REASONING APPROACH BASED ON BELIEF RULE BASE

The calculation and synthesis of activation weights for each rule in the rule base is the core part of the inference system of the belief rule base. The whole process mainly includes two steps: calculate the activation weight, synthesize the rules according to the activation weight. The calculation of the activation weight of each rule in the belief rule base can be regarded as calculating the belief distribution similarity on each attribute and combining their results. Euclidean distance is used to calculate the individual matching degree of the i-th attribute. After converting the input data to have the same belief distribution form as the corresponding attribute, the individual matching degree of the attribute is calculated as:

$$S_i^k = 1 - d_i^k = 1 - \sqrt{\frac{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}{2}}$$

After the individual matching degree of each attribute is calculated, the individual matching degrees of all attributes are aggregated. The aggregation function in the form of conjunctive rules is:

$$\alpha_k = \prod_{i=1}^{T_k} (S_i^k)^{\overline{\delta_i}}, \overline{\delta_i} = \frac{\delta_i}{\max_{j=1,\cdots,T_k} \delta_j}$$

The activation weight of this rule is calculated by the following formula:

$$w_k = \frac{\theta_k \alpha_k}{\sum_{l=1}^{L} \theta_l \alpha_l}$$

Rule weight normalization operation makes all weights satisfy $0 \leq w_k \leq 1, \sum w_k = 1$.

After the rule weight calculation is completed, all the rules are synthesized and the inference result is obtained. First, the belief distribution of the rule is transformed into the corresponding probability quality information:

$$m_{j,k} = w_k \beta_j^k, j = 1, \cdots, N$$

$$m_{D,k} = 1 - \sum_{j=1}^{N} m_{j,k} = 1 - w_k \sum_{j=1}^{N} \beta_j^k$$

$$\overline{m}_{D,k} = 1 - w_k, \quad \widetilde{m}_{D,k} = w_k (1 - \sum_{j=1}^{N} \beta_j^k)$$

$m_{j,k}$ represents the credibility of the k rule on the j result attribute, $\overline{m}_{D,k}$ represents the credibility that the k-th rule is not assigned to any result attribute, $\widetilde{m}_{D,k}$ represents the credibility of the missing reference attribute of the result of the k-th rule, The total uncertainty credibility is given by $m_{D,k} = \overline{m}_{D,k} + \widetilde{m}_{D,k}$. Synthesize according to the credibility information of each rule and get the final belief result of the j-th result attribute:

$$m_j = k[\prod_{i=1}^{L} (m_{j,i} + m_{D,i}) - \prod_{i=1}^{L} m_{D,i}]$$

$$\overline{m}_D = n[\prod_{i=1}^{L} \overline{m}_{D,i}], \quad \widetilde{m}_D = k[\prod_{i=1}^{L} m_{D,i} - \prod_{i=1}^{L} \overline{m}_{H,i}]$$

$$n = [\sum_{j=1}^{N} \prod_{i=1}^{L} (m_{j,i} + m_{D,i}) - (N-1) \prod_{i=1}^{L} m_{D,i}]^{-1}$$

$$\beta_j = \frac{m_j}{1 - \overline{m}_D}, \quad \beta_D = \frac{\widetilde{m}_D}{1 - \overline{m}_D}$$

C. NEW ATTRIBUTE STRUCTURE AND RULE ACTIVATION WEIGHT CALCULATION METHOD

The process of generating rules in the rule base requires artificial setting of candidate reference values for the antecedent attribute information, and when calculating the rule activation weight, if the input attribute information is not in the adjacent area of the rule attribute reference value, the rule cannot be activated. If all the rules in the library are not activated, the reasoning system will fail. In order to solve the above problems, we proposes an improved form of belief

Figure_1.png,Figure_2.png

FIGURE 1. activation weight distributions

rules and corresponding activation weight calculation method as follows:

$$R_k : if(x_1^k, \cdots, x_{T_k}^k)$$

$$then\{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\}, \sum_{i=1}^{N} \beta_i^k \leq 1$$

The simplified belief rule structure can directly use the training data to generate the rule antecedent attribute information without manually setting the candidate reference values of the antecedent attributes. Using antecedent attribute belief distribution similarity as the activation weight calculation method is no longer applicable to the simplified confidence rule form. In order to perform effective weight activation, we uses Gaussian function to calculate the individual matching degree for activation weight calculation. The degree of individual matching of input data $X = (x_1, \cdots, x_(T_k))$ and rule $R_k : if(x_1^k, \cdots, x_{T_k}^k) then\{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\}$ on i-th attribute is calculated using the Gaussian function as:

$$S_i^k = e^{-[a_i^k \times (x_i - x_i^k)]^2}$$

The parameter a_i^k represents the sensitivity of the i-th attribute to the distance at the position x_i^k. When the distance between the rule attribute information and the input data remains unchanged, the larger the parameter a, the closer the attribute matching degree is to 0. The activation weight of a single rule under conjunctive conditions is calculated by the following formula:

$$w_k = \prod_{i=1}^{T_k} S_i^k = e^{-\sum_{i=1}^{T_k} [a_i^k \times (x_i - x_i^k)]^2}$$

Assuming a rule with two attributes x and y located at the origin, the traditional method and Gaussian function method are used to calculate the activation weights. Set the reference candidate values on the x-attributes and y-attributes to be both $[-4, -3, -2, -1, 0, 1, 2, 3, 4]$, and set the distance-sensitive parameter a of each attribute at the origin is 0.5, the two activation weight distributions shown in Figure 1 and Figure 2 can be obtained.

III. GRADIENT DESCENT METHOD PARAMETER TRAINING

wtf

REFERENCES

[1] YANG J B, LIU J, WANG J, et al. "Belief rule-base inference methodology using the evidential reasoning approach-rimer," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 2006, 2(36):266-285.

[2] SUN R. "Robust reasoning: integrating rule-based and similarity-based reasoning," *Artificial Intelligence*, 1995, 2(75):241-295.

[3] DEMPSTER A P. "A generalization of bayesian inference," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1968, 2(30): 205-232.

[4] SHAFER G, SMITH A F M. "A mathematical theory of evidence[J]," *Biometrics*, 1976, 3(32):703.

[5] YOON K, HWANG C L. "Multiple attribute decision making," *Thousand Oaks, CA: Sage Publications*, 1995.

[6] ZADEH L. "Fuzzy sets," *Information and Control*, 1965, 3(8):338-353.

[7] ZHOU Z J, HU C H, YANG J B, et al. "Online updating belief rule based system for pipeline leak detection under expert intervention," *Expert Systems with Applications*, 2009, 4(36):7700-7709.

[8] JIANG J, LI X, JIE ZHOU Z, et al. "Weapon system capability assessment under uncertainty based on the evidential reasoning approach," *Expert Systems with Applications*, 2011.

[9] YANG Y, FU C, CHEN Y W, et al. "A belief rule based expert system for predicting consumer preference in new product development," *Knowledge-Based Systems*, 2016(94):105-113.

[10] YANG J B, LIU J, XU D L, et al. "Optimization models for training belief-rule-based systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 2007, 4(37):569-585.

[11] CHANG R, ZHANG S. "An Algorithm for Training Param eters in Belief Rule-bases Based on Gradient M ethods with Optimization Step Size," *Journal of North China Institute of Water Conservancy and Hydroelectric Power*, 2011, 1(32):154-157.

[12] CHANG R Y, WANG H, YANG J B. "An algorithm for training parameters in belief rule-bases based on the gradient and dichotomy methods," *Systems Engineering*, 2007.

[13] WU W K, YANG L H, FU Y G, et al. "Parameter Training Approach for Belief Rule Base Using the Accelerating of Gradient Algorithm," *Journal of Frontiers of Computer Science and Technology*, 2014, 8(8):989-1001.

[14] SU Q, YANG L H, FU Y G,et al. "Parameter training approach based on variable particle swarm optimization for belief rule base," *Journal of Computer Applications*, 2014, 34(8):2161-2165.

[15] WANG H J, YANG L H, FU Y G H, et al. "Differential Evolutionary Algorithm for Parameter Training of Belief Rule Base under Expert Intervention," *Computer Science*, 2015, 42(5):88-93.

[16] LIU J, MARTINEZ L, CALZADA A C, et al. "A novel beliefrule base representation, generation and its inference methodology," *Knowledge-Based Systems*, 2013(53):129-141.

[17] ROBBINS H, MONRO S. "A stochastic approximation method," *The Annals of Mathematical Statistics*, 1951, 3(22):400-407.

[18] KIWIEL K C. "Convergence and efficiency of subgradient methods for quasiconvex minimization," *Mathematical Programming*, 2001,1(90):1-25.

[19] RUMELHART D E, HINTON G E, WILLIAMS R J. "Learning representations by back-propagating errors," *Nature*, 1986, 6088(323):533-536.

[20] LIN Y Q, FU Y G. "A rule activation method for extended belief rule base based on improved similarity measures," *Journal of University of Science and Technology of China*, 2018, 48(1):21-27.

[21] LIN Y Q, FU Y G, SU Q, et al. "A rule activation method for extended belief rule base with vp-tree and mvp-tree," *Journal of Intelligent and Fuzzy Systems*, 2017, 33(6):3695-3705.

[22] FANG W, GONG X, LIU G, et al. "A balance adjusting approach of extended belief-rule-based system for imbalanced classification problem," *IEEE Access*, 2020, 8(19419049):41201-41212.

[23] SHANNON C E. "A mathematical theory of communication," *Bell System Technical Journal*, 1948, 4(27):623-656.

[24] CALZADA A, LIU J, WANG H, et al. "A new dynamic rule activation method for extended belief rule-based systems," *IEEE Transactions on Knowledge and Data Engineering*, 2015, 4(27):880-894.

[25] YANG L H. "New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis," *Knowledge and Information Systems*, 2018, 60(2):837-878.

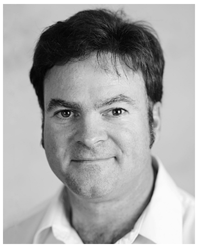
[26] WANG Y. "Parameter learning for an intuitionistic fuzzy belief rulebased systems based on weight and reliability," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2019, 23(2):219-228.

[27] ZHU H. "A minimum centre distance rule activation method for extended belief rule-based classification systems," *Applied Soft Computing*, 2020, 91:106214.

[28] JIA Q. "A novel fault detection model based on atanassov's intervalvalued intuitionistic fuzzy sets, belief rule base and evidential reasoning," *IEEE Access*, 2020, 8:4551-4567.

FIRST A. AUTHOR (M'76–SM'81–F'87) and all authors may include biographies. Biographies are often not included in conference-related papers. This author became a Member (M) of IEEE in 1976, a Senior Member (SM) in 1981, and a Fellow (F) in 1987. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state, and country, and year the degree was earned. The author's major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (publisher name, year) similar to a reference. Current and previous research interests end the paragraph. The third paragraph begins with the author's title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, it should be of good quality, and professional-looking. Following are two examples of an author's biography.



SECOND B. AUTHOR was born in Greenwich Village, New York, NY, USA in 1977. He received the B.S. and M.S. degrees in aerospace engineering from the University of Virginia, Charlottesville, in 2001 and the Ph.D. degree in mechanical engineering from Drexel University, Philadelphia, PA, in 2008.

From 2001 to 2004, he was a Research Assistant with the Princeton Plasma Physics Laboratory. Since 2009, he has been an Assistant Professor with the Mechanical Engineering Department, Texas A&M University, College Station. He is the author of three books, more than 150 articles, and more than 70 inventions. His research interests include high-pressure and high-density nonthermal plasma discharge processes and applications, microscale plasma discharges, discharges in liquids, spectroscopic diagnostics, plasma propulsion, and innovation plasma applications. He is an Associate Editor of the journal *Earth, Moon, Planets*, and holds two patents.

Dr. Author was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award in 2011.

THIRD C. AUTHOR, JR. (M'87) received the B.S. degree in mechanical engineering from National Chung Cheng University, Chiayi, Taiwan, in 2004 and the M.S. degree in mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2006. He is currently pursuing the Ph.D. degree in mechanical engineering at Texas A&M University, College Station, TX, USA.

From 2008 to 2009, he was a Research Assistant with the Institute of Physics, Academia Sinica, Tapei, Taiwan. His research interest includes the development of surface processing and biological/medical treatment techniques using nonthermal atmospheric pressure plasmas, fundamental study of plasma sources, and fabrication of micro- or nanostructured surfaces.

Mr. Author's awards and honors include the Frew Fellowship (Australian Academy of Science), the I. I. Rabi Prize (APS), the European Frequency and Time Forum Award, the Carl Zeiss Research Award, the William F. Meggers Award and the Adolph Lomb Medal (OSA).