

REGULARIZATION METHOD FOR RULE REDUCTION IN BELIEF RULE-BASED SYSTEM

Yu Guan

College of Mathematics and Computer Science,
Fuzhou University, Fuzhou, China

ABSTRACT

Belief rule-based inference system introduces a belief distribution structure into the conventional rule-based system, which can effectively synthesize incomplete and fuzzy information. In order to optimize reasoning efficiency and reduce redundant rules, this paper proposes a rule reduction method based on regularization. This method controls the distribution of rules by setting corresponding regularization penalties in different learning steps and reduces redundant rules. This paper first proposes the use of the Gaussian membership function to optimize the structure and activation process of the belief rule base, and the corresponding regularization penalty construction method. Then, a step-by-step training method is used to set a different objective function for each step to control the distribution of belief rules, and a reduction threshold is set according to the distribution information of the belief rule base to perform rule reduction. Two experiments will be conducted based on the synthetic classification data set and the benchmark classification data set to verify the performance of the reduced belief rule base.

KEYWORDS

Knowledge-based system, Belief rule base, Regularization method, Rule reduction.

1. INTRODUCTION

The inference system of the belief rule base proposed by Yang et al. with a belief distribution structure and evidential reasoning method is based on the research results of D-S evidence theory, fuzzy theory, and generative IF-THEN rules. Belief rule-based inference system can effectively synthesize the missing, fuzzy, and uncertain parts of the input information. In the inference process of the belief rule base, the attribute weight, rule weight, belief distribution, and other parameters in the system directly affect the accuracy of the final inference prediction result. To improve the inference accuracy of the belief rule base, Yang et al. proposed a parameter optimization model of the belief rule base. Later researches also proposed a series of belief rule base parameter optimization models using different machine learning algorithms. The early belief rule base can only construct rules based on the specific domain knowledge of the human expert, and cannot construct a reasoning system containing a large number of rules. The extended belief rule-based inference system uses the training data set to construct the rule base based on the data-driven concept.

This paper proposes the optimized belief rule base further simplifies the belief distribution structure of the belief on the basis of the extended belief rule base system, further improves the efficiency of constructing belief rules through training data sets, and reduces the complexity of rule storage and operation. Then a parameter training method based on regularization is proposed to reduce redundant rules. The rest of this paper is organized as follows: In Section II, we

Natarajan Meghanathan et al. (Eds): CSE, BMLI, EDU, AI & FL, CCNET - 2020
pp. 51-61, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101705

reviewed the basic structure and parameter training model of the belief rule base. Then Section III proposes the optimized belief rule structure, evidence reasoning method, and parameter training method based on regularization. The two experiments in Section IV verify the performance of the reduced belief rule base system, and Section V concludes the paper.

2. OVERVIEW OF BELIEF RULE BASE INFERENCE SYSTEM

The construction of the belief rule base is based on the IF-THEN generative rules and the belief distribution framework, and the professional domain knowledge is obtained through expert setting or other methods. The inference engine uses the D-S evidence synthesis theory to synthesize the conclusions of different rules to obtain the final reasoning result. This section will specifically introduce the structure of the belief rule base and the process of the evidential reasoning of the belief rule base.

2.1. Belief Rule Base

The belief rule base proposed by Yang et al. is based on the traditional IF-THEN generative rules, and introduces the structure of belief distribution on the result attribute, and introduces the weight of the antecedent attributes and the rule weight, which can effectively express uncertain information. The structure of the belief rule base has L rules, T attributes, and N results is as follows:

$$R_k: \text{if } \{x_1 \text{ is } A_1^k \wedge \cdots \wedge x_T \text{ is } A_T^k\} \text{ then } \{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\} \quad (1)$$

with rule weight $\theta_k, k = 1, \cdots, L$ and attribute weight $\delta_1, \cdots, \delta_T$

(x_1, \cdots, x_T) belong to a certain reference candidate of their antecedent attributes, for any A_i^k satisfies $A_i^k \in \{A_{i,1}, \cdots, A_{i,J_i}\}$. In the extended belief rule-based inference system, the belief distribution structure is further introduced into the antecedent attributes, which improves the model's ability to express fuzziness and incomplete information. The k th rule in the extended belief rule base can be expressed as:

$$R_k \text{ if: } \{[(A_{11}^k, \alpha_{11}^k), \cdots, (A_{1J_1}^k, \alpha_{1J_1}^k)] \wedge \cdots \wedge [(A_{T1}^k, \alpha_{T1}^k), \cdots, (A_{TJ_T}^k, \alpha_{TJ_T}^k)]\} \\ \text{then } \{(D_1, \beta_1^k), \cdots, (D_N, \beta_N^k)\} \quad (2)$$

with rule weight $\theta_k, k = 1, \cdots, L$ and attribute weight $\delta_1, \cdots, \delta_T$

The rules in the extended belief rule base can be directly generated from the training data set. For the input data, convert the t th attribute parameter to construct the t th antecedent attribute of the corresponding rule in the form of belief distribution using the corresponding reference values of candidate attribute $\{(A_{i1}, \gamma_{i1}), \cdots, (A_{iJ_i}, \gamma_{iJ_i})\}$:

$$\alpha_{ij}^k = \frac{\gamma_{i(j+1)} - x_i^k}{\gamma_{i(j+1)} - \gamma_{ij}}, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)} \\ \alpha_{i(j+1)}^k = 1 - \alpha_{ij}^k, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)} \\ \alpha_{it}^k = 0, t = 1, \cdots, (j-1), (j+2), \cdots, J_i \quad (3)$$

2.2. Reasoning Method using Evidential Reasoning

The belief rule-based reasoning system uses the evidential reasoning method to synthesize rule results. The inference process consists of the following steps in sequence:

1) belief rule activation weight calculation

For the component of input on any antecedent attribute, convert it to the belief distribution on the corresponding attribute, and the method is as follows:

$$\begin{aligned}\alpha_{ij}^k &= \frac{\gamma_{i(j+1)} - x_i}{\gamma_{i(j+1)} - \gamma_{ij}}, \gamma_{ij} \leq x_i \leq \gamma_{i(j+1)} \\ \alpha_{i(j+1)} &= 1 - \alpha_{ij}, \gamma_{ij} \leq x_i \leq \gamma_{i(j+1)} \\ \alpha_{it} &= 0, t = 1, \dots, (j-1), (j+2), \dots, J_i\end{aligned}\quad (4)$$

Using the belief distribution after input conversion, the individual matching degree of the k th rule on i th antecedent attribute is calculated as:

$$S_i^k = 1 - d_i^k = 1 - \sqrt{\frac{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}{2}} \quad (5)$$

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:

$$\alpha_k = \prod_{i=1}^{T_k} (S_i^k)^{\delta_i}, \delta_i = \frac{\delta_i}{\max_{j=1, \dots, T_k} \delta_j} \quad (6)$$

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

$$w_k = \frac{\theta_k \alpha_k}{\sum_{l=1}^L \theta_l \alpha_l} \quad (7)$$

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

2) evidential reasoning of belief rule base

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 mass information:

$$\begin{aligned}
m_{j,k} &= w_k \beta_j^k, j = 1, \dots, N \\
m_{D,k} &= 1 - \sum_{j=1}^N m_{j,k} = 1 - w_k \sum_{j=1}^N \beta_j^k \\
\bar{m}_{D,k} &= 1 - w_k \\
\tilde{m}_{D,k} &= w_k \left(1 - \sum_{j=1}^N \beta_j^k\right)
\end{aligned} \tag{8}$$

where $m_{j,k}$ represents the credibility of the k th rule on the j th consequent attribute, $\bar{m}_{D,k}$ represents the credibility that the k th rule is not assigned to any consequent attribute, and $\tilde{m}_{D,k}$ represents the credibility of the missing reference attribute of the k th rule. The total uncertainty credibility is given by $m_{D,k} = \bar{m}_{D,k} + \tilde{m}_{D,k}$.

Synthesize the credibility information of all rules and obtain the final belief result of each consequent attribute:

$$\begin{aligned}
m_j &= k \left[\prod_{i=1}^L (m_{j,i} + m_{D,i}) - \prod_{i=1}^L m_{D,i} \right], j = 1, \dots, N \\
\bar{m}_D &= n \left[\prod_{i=1}^L \bar{m}_{D,i} \right] \quad \text{and} \quad \tilde{m}_D = k \left[\prod_{i=1}^L m_{D,i} - \prod_{i=1}^L \bar{m}_{H,i} \right] \\
k &= \left[\sum_{j=1}^N \prod_{i=1}^L (m_{j,i} + m_{D,i}) - (N-1) \prod_{i=1}^L m_{D,i} \right]^{-1} \\
\beta_j &= \frac{m_j}{1 - \bar{m}_D}, j = 1, \dots, N \quad \text{and} \quad \beta_D = \frac{\tilde{m}_D}{1 - \bar{m}_D}
\end{aligned} \tag{9}$$

3. BELIEF RULE BASE STRUCTURE OPTIMIZATION AND REGULARIZATION METHOD

This section optimizes the traditional belief rule base structure by simplifying the belief structure of antecedent attributes and introduces the Gaussian membership function to optimize the calculation of activation weights. This section also proposes a group-level evidential reasoning method to avoid reasoning failure when there are too many rules. For the optimized inference system of the belief rule base, the regularization method is used to restrict and select different parameters in the belief rule base step by step during the training process, and the rules are further screened and reduced according to the rule parameters after training.

3.1. Structural Optimization of Belief Rule

The rule structure used by the conventional belief rule-based inference system is based on the belief distribution form. When constructing the belief rule and inferring the input data, it is necessary to convert the data into the corresponding belief distribution form, which requires additional computing and storage resources. It is also necessary to set the attribute candidate reference values in advance, and the empirical knowledge of human experts is required. The

unreasonable setting of the number of candidates and values will reduce the accuracy of the system's reasoning.

Using the Euclidean distance method to calculate the attribute similarity may calculate abnormal activation weights, which may cause the inference system to fail due to the rule zero activation problem. To prevent the rule zero activation problem, it is necessary to construct more rules to cover all the possibilities, which causes the explosion of the number of rules caused by the increase in the number of attributes.

Because of the above shortcomings, this section first optimizes the rule structure by simplifying the belief distribution structure to avoid using the activation weight calculation method based on the similarity of the belief distribution, avoiding the problem of reasoning failure that it may cause, and simplifying the complexity of the rule construction. For training data $X_k = (x_1^k, \dots, x_T^k)$, the corresponding belief rule that directly simplifies the belief distribution structure of the antecedent attributes is as follows:

$$R_k: \text{if } \{x_1^k \wedge \dots \wedge x_T^k\} \text{ then } \{(D_1, \beta_1^k), \dots, (D_N, \beta_N^k)\} \quad (10)$$

with rule weight $\theta_k, k = 1, \dots, L$ and attribute weight for each rule $\delta_1^k, \dots, \delta_T^k$

The simplified belief rule directly uses the attribute information corresponding to the original data to construct without the conversion of the belief distribution, avoiding additional computing and storage resources. The optimized belief rule structure cannot calculate the individual matching degree of the rule attributes by calculating the Euclidean distance of the belief distribution. This section proposes to use the Gaussian membership function to optimize the activation weight calculation process. The Gaussian membership function is a function in fuzzy theory that calculates the degree to which a specified element belongs to a specific set. Its form is as follows:

$$\text{gaussianmf}(x; \delta, c) = e^{-\frac{(x-c)^2}{\delta^2}} \quad (11)$$

For the input data $X = (x_1, \dots, x_T)$, the individual matching degree on the i th attribute of k th rule is calculated using the Gaussian membership function:

$$S_i^k = e^{-\frac{[\delta_i^k \times (x_i - x_i^k)]^2}{\theta_k^2}} \quad (12)$$

The individual matching degree is determined by the distance from the input attribute information to the corresponding rule attribute, the rule attribute weight parameter, and the rule weight parameter. Unlike the conventional belief rule-based inference system that uses uniform attribute weights, each rule has its own attribute weight parameter, which can achieve finer rule activation granularity. Under the confidence rule setting of the conjunctive relationship, the final activation weight of this rule is calculated by the following formula:

$$w_k = \prod_{i=1}^T S_i^k = e^{-\frac{\sum_{i=1}^T [\delta_i^k \times (x_i - x_i^k)]^2}{\theta_k^2}} \quad (13)$$

The calculated activation weight is within the range of $(0, 1]$ without the need for weight normalization, which simplifies the reasoning process. There will be no rule activation value zero, which avoids the potential rule zero activation problem and improves the robustness of the inference system.

3.2. Parameter Training of Regularization Method

This section uses the step-by-step regularization method. First, the regularization penalty of the rule antecedent attributes and attribute weights is introduced in the training process to make the belief rule distribution of the inference system more widely representative. Then the antecedent attributes and attribute weights are fixed, and the regularization penalty on the rule weight is introduced. The distribution of the activation weight of the belief rule base is restricted through training, and the rule weight is used to measure the importance of the corresponding rule after training. After training, the rule weight is used to determine whether to retain the corresponding rule, and the rule reduction of the belief rule base is realized.

1) complexity expression of belief rule base

According to the belief rule structure in equation (10), both the rule antecedent attributes and the corresponding attribute weights need to be restricted within a certain range. The restriction of antecedent attributes makes the rule not far away from the data distribution, which makes the rule redundant and useless. The restriction of the attribute weights makes the activation of the rule always represent the distribution of a part of the data, avoiding too high attribute weight to make the rule fit noisy data. The objective function can be expressed when using the L_2 regularization method to construct the belief rule base penalty:

$$obj_1 = \min\{loss + \lambda_x \|x\|_2^2 + \lambda_\delta \|\delta\|_2^2\} \quad (14)$$

The regularization coefficients λ_x and λ_δ are used to control the degree of the penalty of the model, to prevent the model penalty from being too large or too small from affecting the final inference accuracy. The calculation of the rule activation weight in equation (13) is also affected by the rule weight. When using the objective function, the rule weight needs to be fixed in advance to avoid model penalty failure. When using the L_1 regularization method to construct the objective function that includes the model penalty, according to the belief rule structure in (10) and the calculation method of activation weight in equation (13), The penalty of rule weight can effectively restrict the activation range of the rule, so that the activation range of the redundant rule that activates the same distributed data is limited and reduced. By comparing the weight parameters of the rules after training, important rules can be selected for rule reduction, and the corresponding objective function can be expressed as:

$$obj_2 = \min\{loss + \lambda_\theta \|\delta\|_1\} \quad (15)$$

In the same activation weight calculation process, antecedent attributes and attribute weights are involved. The use of the objective function requires fixed antecedent attributes and attribute weights.

2) rule reduction of step-by-step regularization training method

Combining the newly proposed belief rule structure, regularization penalty construction method, and group-level rule evidential reasoning method, This chapter proposes a step-by-step method of

parameter training with rule reduction capabilities. Each step fixes different parameters and sets a specific regularization penalty to filter out representative belief rules. The specific steps of the step-by-step regularization parameter training method are as follows:

Algorithm 1: Regularization Parameter Training Method

Input: the attributes of belief rule base and regularization coefficients $x, \delta, \theta, \lambda_x, \lambda_\delta, \lambda_\theta$, the training data set D and obj_1, obj_2 function and model $loss = Loss(D; x, \delta, \theta)$

Output: Reduced belief rule base

- 1: Fixed rule weight δ , set obj_1 as the objective function with $\lambda_x, \lambda_\delta$, Use training set D for training
- 2: Fixed antecedent attribute x and attribute weights δ , Set obj_2 as the objective function with λ_δ , Use training set D for another training
- 3: Set reduction threshold ts based on rule weight after training and reduce rules with rule weight less than ts
- 4: Set $loss$ as the objective function, Use training set D to train the reduced belief rule base

Return the reduced belief rule base attributes

4. EXPERIMENTS

This section first uses a synthetic binary classification data set containing two numerical attributes to calculate the activation weight distribution of the belief rule base after each training step is completed and verify the effectiveness of the regularization method to limit the activation range of the rule. Then the UCI public classification data set is used to compare the inference performance of the belief rule base after reduction under different reduction threshold parameter settings.

4.1. Synthetic Binary Data Set Experiment

This section selects the artificially generated binary data set with two numerical attribute variables used in [1], which contains a total of 250 data, with 125 positive and negative samples each. To facilitate the setting of the attribute weights, the two numerical attribute variables are standardized before the experiment. The standardized data distribution is shown in Figure 1:

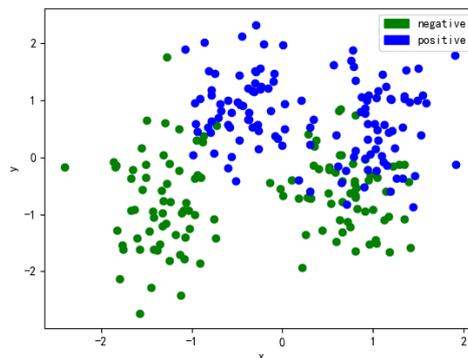


Figure 1. Synthetic binary data set distribution

Four samples are randomly selected from the standardized positive and negative training samples and the corresponding belief rules are constructed. The attribute weights and rule weights of all

confidence rules are initialized to 1.0. Figure 2 shows the activation weight distribution of the belief rule corresponding to the negative and positive samples.

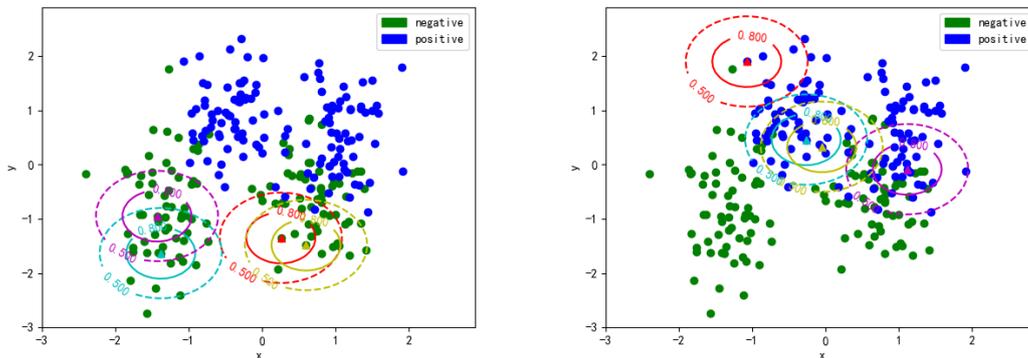


Figure 2. The activation weight distribution of the belief rule generated by the negative and positive sample

In the first step of regularization training, the rule weight is fixed and the penalty coefficients of the antecedent attributes and attribute weights are both set to 0.001. Use cross-entropy as the classification loss function and gradient method as the parameter training method.

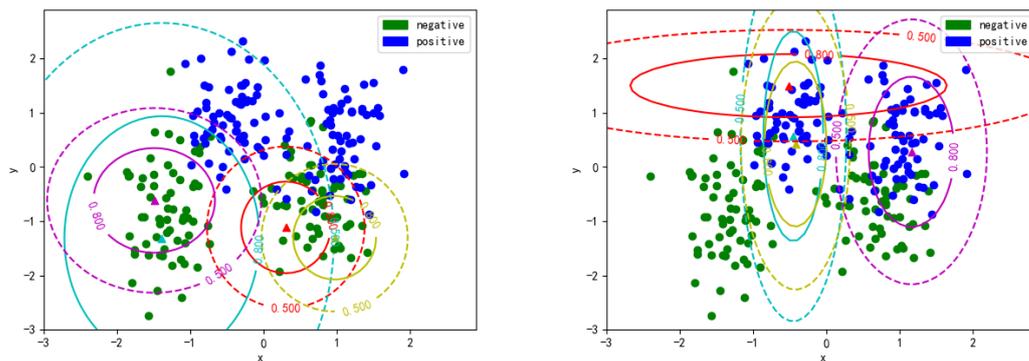


Figure 3. The activation weight distribution of the belief rule generated by the negative and positive sample after the first step training

Figure 3 shows the activation weight distribution of the belief rule corresponding to the positive and negative samples after the first training. It can be found that the regularization method makes the activation weight distribution of each rule approach each other, and the antecedent attribute distribution and attribute weight of the belief rule are also more similar. It provides favourable conditions for further rule reduction.

In the second step of regularization training, fix the antecedent attributes and attribute weights, and set the rule weight penalty coefficient to 0.001. Use the same parameter optimization method for parameter training. Figure 4 shows the rule activation weight distribution corresponding to the positive and negative samples after the training. It can be found that the activation range of the rules with similar activation weight distribution is reduced to negligible, and the remaining rules with a larger activation range cover the activation area of the rule with the reduced activation range.

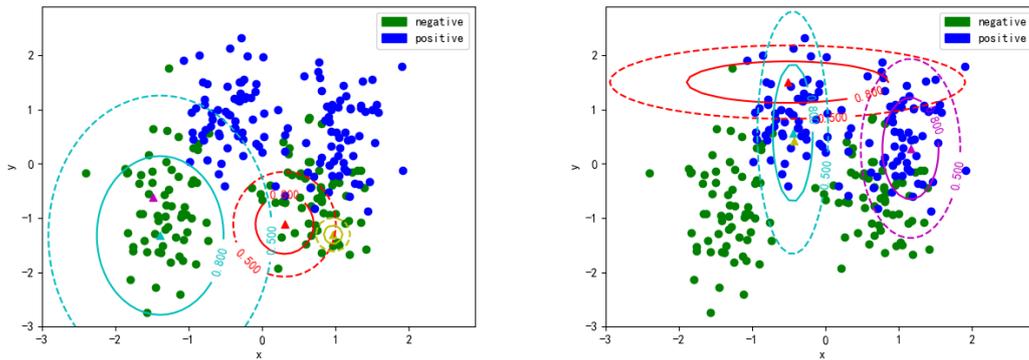


Figure 4. The activation weight distribution of the belief rule generated by the negative and positive sample after the second step training

The rule weights corresponding to the positive and negative samples after regularization training in the second step are listed in Table 1. The maximum rule weight is 0.2287, and the threshold is set to 50% of the maximum rule weight, which is 0.1143 for rule reduction. Rules 1 and 2 for positive samples and rule 4 for negative samples are retained.

Table 1. Rule weight after the second step training

No.	Rules(from positive samples)				Rules(from negative samples)			
	1	2	3	4	1	2	3	4
Weight	0.13530	0.1617	0.0040	0.0879	0.0997	0.6780	0.0327	0.2287

The reduced belief rule base contains three rules, and the third step of training is performed to adjust the reduced belief rule base. The activation weight distribution of the three rules and classification contour maps after training are shown in Figure 5.

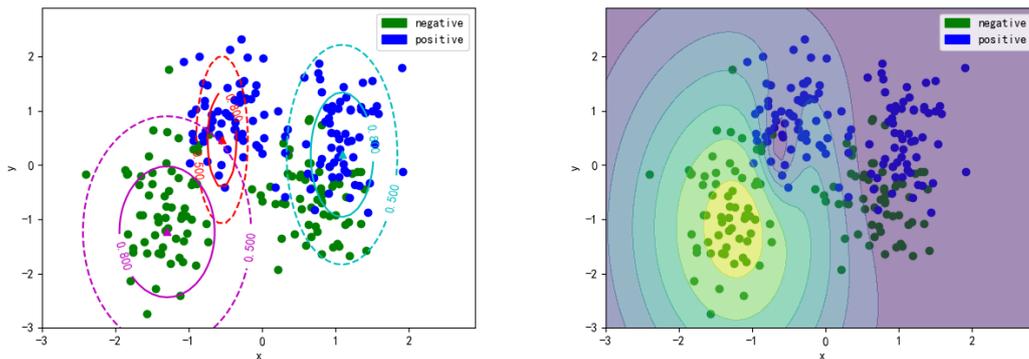


Figure 5. The activation weight distribution of the three rules and classification contour maps after training

4.2. Benchmark Data Sets Experiment

This section uses four UCI classification data sets to verify the reduction performance of the regularization training method on the belief rule base. Table 2 lists the detailed information of the data. Each data set repeats ten independent ten-fold cross-validation experiments to obtain the final results.

Unified data standardization before the experiment, using all training sets to construct the belief rule base. The attribute weight and rule weight are both set to 1.0, and the penalty coefficient of each step of the regularization method is set to 0.001. The experiment compares the reduction size and inference accuracy of the belief rule base under different reduction threshold settings of 10%, 30%, 50%, 70% and compare the number of rules and inference accuracy with the compact belief rule-based classification system using evidence clustering(CBRBCS)[5].

Table 3. Details of the classification datasets

Dataset	#Instances	#Features	#Classes
Iris	150	5	3
Wine	178	14	3
Ecoli	336	8	8
Glass	214	10	6

Table 3 lists the number of rules and the corresponding inference accuracy after the reduction of belief rule base using regularization training method (BRB-R) and the CBRBCS inference system after regularization reduction training on the Iris, Wine, Ecoli, and Glass datasets. Observing the data in the table, we can find that as the reduction threshold increases, the method in this chapter has no significant decrease in inference accuracy and can reduce more rules.

Table 4. Comparison of the results for classification datasets

Method	Aspects	Iris	Wine	Ecoli	Glass
CBRBCS	Accuracy	93.33	94.80	82.62	68.15
	Reduction rate	21.43%	86.86%	17.78%	45.00%
BRB-R(10%)	Accuracy	95.03	94.84	84.94	71.61
	Reduction rate	72.15%	77.63%	72.30%	68.79%
BRB-R(30%)	Accuracy	94.83	95.03	85.12	71.37
	Reduction rate	84.88%	87.81%	69.57%	70.52%
BRB-R(50%)	Accuracy	93.39	95.10	85.64	70.34
	Reduction rate	90.88%	93.56%	78.26%	74.06%
BRB-R(70%)	Accuracy	93.63	93.14	86.13	71.81
	Reduction rate	99.57%	96.56%	87.37%	84.16%

5. CONCLUSIONS

This paper proposes a step-by-step regularization parameter training method by constructing the complexity penalty of the reasoning system based on belief rules. Each step of the training is fixed with different parameters, and different penalties are selected to achieve data fitting and rule reduction. The experimental results show that compared with other reduction methods based on belief rule base, this method has a higher simplification rate and higher inference accuracy.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Fujian Province, China (No. 2019J01647), and the Industrial Internet Innovation and Development Project of the Ministry of Industry and Information Technology of China (No. TC19083WB)

REFERENCES

- [1] B.D. Ripley "Pattern Recognition and Neural Networks," Cambridge University Press, 1996.
- [2] J. B. Yang, J. Liu, J. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 36, no. 2, pp. 266-285, 2006.
- [3] J. Liu, L. Martinez, A. C. Calzada, "A novel belief rule base representation, generation and its inference methodology," Knowledge-Based Systems, vol. 53, pp. 129-141, 2013.
- [4] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net," Journal of the royal statistical society: series B (statistical methodology), vol. 67, no. 2, pp. 301-320, 2005.
- [5] L. Jiao, X. Geng, Q. Pan. "Compact belief rule base learning for classification with evidential clustering," Entropy, vol. 21, no. 5, pp. 443, 2019.
- [6] A. P. Dempster, "A generalization of bayesian inference," Journal of the Royal Statistical Society: Series B (Methodological), vol. 30, no. 2, pp. 205-232, 1968.
- [7] G. Shafer, A. F. M. Smith, "A mathematical theory of evidence," Biometrics, vol. 32, no. 3, pp. 703, 1976.
- [8] L. Jiao, T. Denoeux, and Q. Pan, "A hybrid belief rule-based classification system based on uncertain training data and expert knowledge," IEEE Trans. Syst., Man, Cybern. Syst., vol. 46, no. 12, pp. 1711–1723, 2016.
- [9] L. Jiao, Q. Pan, T. Denoeux, Y. Liang, and X. Feng, "Belief rule based classification system: Extension of FRBCS in belief functions framework," Inform. Sciences, vol. 309, no. 1, pp. 26–49, 2015.

AUTHOR

Yu Guan received the B.S. degree from Fuzhou University, in 2018, where he is currently pursuing the master's degree. His research interests include intelligent decision technology, rule-based inference, big data analysis, and machine learning.

