

Research Article

Optimal Granularity Selection for Indoor Localization Detection with Wireless IoT Networks

Feng Cao,¹ Jing Zhang,² Deyu Li ,^{1,3} Yuhua Qian,^{1,3,4} Chao Tang ,⁵ Xialei Zhang,¹ and Zhiguo Hu¹

¹School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

²Department of Math, Taiyuan College, Taiyuan 030006, China

³Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan 030006, China

⁴Institute of Big Data and Industry, Shanxi University, Taiyuan 030006, China

⁵Department of Computer and Science Technology, Hefei College, Hefei 230601, China

Correspondence should be addressed to Deyu Li; lidysxu@163.com

Received 18 May 2022; Revised 26 June 2022; Accepted 25 July 2022; Published 10 August 2022

Academic Editor: Changyan Yi

Copyright © 2022 Feng Cao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Indoor localization detection acts as an important issue and has wide applications with wireless Internet of Things (IoT) networks. In recent years, the WiFi-based localization by using the latest artificial intelligence methods for improving the detection accuracy has attracted attention of many researchers. Granular computing is a newly emerged computing paradigm in artificial intelligence, which focuses on the structured thinking based on multiple levels of granularity. Thus, we introduce granular computing approaches to the task of wireless indoor localization detection, and a novel heuristic data discretization method is proposed based on the binary ant colony optimization and rough set (BACORS) for the selection of optimal granularity. For BACORS, the global optimal cut point set is searched based on the binary ant colony optimization to simultaneously discretize multiple attributes. Meanwhile, the accuracy of approximation classifications coined from rough sets is used to determine the consistent of multiple attribute data. To validate the effectiveness of BACORS, it is applied to a wireless indoor localization data set, and the experimental results indicate that it has promising performance.

1. Introduction

With the rapid development of wireless indoor positioning systems and techniques in Internet of Things (IoT), the application of indoor localization detection is more and more extensive [1]. To name a few, one can consider detecting the location of criminals in a bounded area, detecting the location of vehicles in a large underground garage, and finding products stored in a warehouse [2]. In recent years, the indoor localization detection based on WiFi signal strength has received great attention due to its widely available and low cost. In existing literatures, many methods have been proposed by considering positioning as a classification problem, and lots of machine learning techniques have been used to improve the prediction accuracy of localization detection

[3, 4]. For instance, Pei et al. [5] presented a motion recognition-assisted wireless positioning method by using a least square-support vector machines (LS-SVM) to detect common motion states used during indoor navigation. Wang et al. [6] studied a deep network with three hidden layers pretrained with the greedy learning algorithm for location prediction. Zhang et al. [7] explored a novel wireless positioning method by using a four-layer neural network structure pretrained by stacked denoising autoencoder (SDA) to extract features from massive widely fluctuating WiFi data. As received WiFi signal strength is vulnerable to various environments, Zhang et al. [8] proposed a multi-information fusion algorithm to improve the accuracy of indoor location. Rohra et al. [2] used neural network to detect the location of the users in an indoor environment

based on WiFi signal strength received from various routers. To obtain a higher accuracy, fuzzy hybrid of particle swarm optimization and gravitational search algorithm is used to train the weights. Bhatti et al. [9] put forth a technique named iF_Ensemble by using the combination of isolation forest (IForest), support vector machine (SVM), K nearest neighbor (KNN), and random forest (RF) to effectively detect the outlier of users in a WiFi indoor localization environment.

To overcome the drawback that accuracy of indoor localization detection based on WiFi signal strength is easily disturbed by diverse environments, researchers devote to investigating novel detection methods for improving the accuracy. Granular computing has emerged as one of the fastest growing computing paradigms on formation, processing, and communicating information granules in the artificial intelligence field [10], which has been widely applied in many realistic areas, such as private data protection [11], medical diagnose [12], image segmentation [13], and web services recommendation [14]. Thus, it is meaningful to apply granular computing methods to indoor localization detection for improving the detection accuracy.

For granular computing, the construction of information granules and computation with granules are two basic issues. The selection of appropriate granularity levels plays an important role to address application problems. Optimal granularity selection has become one of the hotspots. Discretization is an essential data preprocessing procedure, which has been used to obtain the equivalence-based information granules and improve the performances of granular computing methods. In the past decades, the researches on discretization-based optimal granularity selection have been received significant attention, and many discretization methods have been developed [15, 16]. Those methods can be classified with different taxonomies. Liu et al. [17] summarized the commonly used discretization methods from different taxonomies in detail. Unsupervised and supervised ones are the most commonly used taxonomies among them. Unsupervised discretization methods do not consider class labels during the discretization process. They are generally easy to implement and own low time complexity, whereas supervised discretization methods relate class labels to the discretization process. Therefore, the research work of discretization mainly focuses on the improvement of supervised discretization methods.

Supervised discretization methods mainly include entropy-based, χ^2 measure-based, and dependency-based methods. The main differences of those methods are the choice approaches of optimal cut point sets. Entropy-based discretization methods find the optimal cut point set by using different kinds of entropies to evaluate the importance of cut point sets. MDLP (minimal description length principle) is one of the most common methods among them [18–20], which uses the class information entropy as the measure to evaluate importance of cut points during the discretization process. For a given continuous attribute, the cut point which minimizes the entropy function over all candidate cut points is selected as the optimal cut point. This method can then be

applied recursively to both of the intervals induced by the optimal cut point until the stopping condition MDLP defined by Fayyad and Irani is achieved. χ^2 measure-based discretization methods are a series of supervised discretization methods which evaluate the importance of cut points based on χ^2 measure. ChiMerge and Chi2 are two typical χ^2 measure-based discretization methods. The ChiMerge merges the adjacent pair intervals with lowest χ^2 value until a termination condition is met [21]. One drawback of ChiMerge is that it requires to specify the significant level α when computing the χ^2 value. Nevertheless, too big or too small of an α will over- or under-discretize continuous data, and it is hard to find an appropriate α for ChiMerge. The Chi2 stems from ChiMerge and an autoversion of ChiMerge [22]. Dependency-based discretization methods evaluate the importance of cut points based on the dependency between continuous data and class labels calculated according to contingency table. They mainly include CAIM and CACC. CAIM (class attribute interdependence maximum) defined the interdependency between class labels and the discretization scheme of a continuous attribute. It maximizes the class attribute interdependence and generates a possibly minimal number of cut points [23]. This method does not need to predefine the number of intervals and can always achieve better discretization results. The drawback of CAIM is that the number of intervals generated is very close to the number of target classes. CACC (class-attribute contingency coefficient) is the improvement method of CAIM and overcomes the drawbacks of CAIM. CACC defines contingency coefficient and maximizes it in the process of discretization instead of making class attribute interdependence maximum in CAIM [24].

For the commonly used supervised discretization methods, most of them are local methods performing on a single attribute. They independently generate intervals on each attribute without taking into account the interdependence among attributes. Compared to local discretization methods, global discretization methods discretize multiple continuous attributes simultaneously. The discretization process can be defined as a problem of searching for an optimal cut point set of multiple continuous attributes. Global discretization should obtain less cut points and have better discretization results than local discretization. However, it has been proved by Nguyen and Skowron that the optimization problem to find the optimal cut point set of multiple continuous attributes is NP-hard [25]. Therefore, the commonly used discretization methods are basically heuristic methods by using efficient heuristics to return suboptimal cut point set.

To obtain the optimal equivalence-based information granules, this article proposes a new global heuristic supervised discretization method based on binary ant colony optimization and rough set. It uses binary ant colony optimization to search the optimal least cut point set of multiple attributes. Meanwhile, this method tries to keep the indiscernibility relation defined with rough set unchanged for maintaining the consistent of multiple continuous attributes data. The experiment applying the proposed method to the wireless indoor localization data set indicates that it is effective for improving the prediction accuracy of indoor localization detection.

2. Related Works

2.1. Discretization. Discretization problem investigates how to partition continuous attributes into several discrete intervals with the selected optimal cut points. Let A_i be a continuous attribute of a data set with multiple continuous attributes A and the cut point c be a value within the range $[V_{\min}, V_{\max}]$ of A_i . A cut point set $C = \{c_1, c_2, \dots, c_m | c_i \in [V_{\min}, V_{\max}]\}$ is used to partition A_i into $m + 1$ intervals representing with $m + 1$ different labels during the discretization process. The essential problem of discretization is the selection of optimal cut point set from the candidate cut point set. For continuous attribute A_i , the value set is $V_i = \{a_i^1, a_i^2, \dots, a_i^n\}$ after sorting in ascending order and removing duplicate values, where $a_i^1 < a_i^2 < \dots < a_i^n$. The candidate cut point set of A_i can be defined as $C_{A_i}^{\text{candidate}} = \{c_i^j | c_i^j = (a_i^j + a_i^{j+1})/2, j = 1, \dots, n - 1, \text{ and the candidate cut point set of } A \text{ is}$

$$C_A^{\text{candidate}} = \bigcup_{i=1}^{|A|} C_i^{\text{candidate}}. \quad (1)$$

For the available discretization methods, the main difference is the selection method used to find the optimal cut point set. Local and global discretization methods search the optimal cut point set of single attribute and multiply attributes, respectively. For supervised local or global discretization methods, the optimal cut point set is determined by taking into account the relationship between attributes and class labels.

A discretization process mainly include four steps: (1) sort the continuous values of the attribute and remove the duplicate values, (2) calculate the candidate cut point set of the attribute, (3) search the optimal cut point set, and (4) use the optimal cut point set to discretize the continuous attribute into discrete attributes.

2.2. The Rough Set Theory. The rough set theory proposed by Pawlak et al. is an extension of the classic set theory and proved to be very useful in dealing with inconsistency problems [26]. Since rough sets are useful in analyzing vagueness and uncertain data, it has been widely applied in uncertainty data mining and knowledge discovery. Rough set operates on information system $S = \langle U, CUD, V, f \rangle$, where U is a finite nonempty set of objects, C is a finite nonempty set of condition attributes, D is a finite nonempty set of decision attributes, $V = \bigcup_{a \in CUD} V_a$ is a nonempty set of values of all attributes, and f is an information function such $f(x, a) \in V_a$ for each $a \in A$ and $x \in U$. To each nonempty subset of attributes $A \subseteq C$, an indiscernibility relation can be defined as

$$\text{IND}(A) = \{(x, y) \in U \times U | \forall a \in A f(a, x) = f(a, y)\}. \quad (2)$$

$\text{IND}(A)$ is called the A -indiscernibility relation. The family of all the equivalence classes of the relation $\text{IND}(A)$ containing an object $x \in U$ is denoted as $[x]_A$.

Let S be an information system, $X \subseteq U$ be nonempty subset, and $A \subseteq C$. The lower and upper approximations of X are defined as follows:

$$A_-(X) = \{x \in U | [x]_A \subseteq X\}, \quad (3)$$

$$A^-(X) = \{x \in U | [x]_A \cap X \neq \emptyset\}, \quad (4)$$

$A_-(X)$ is an object set whose objects belonging to the equivalence classes generated by the indiscernibility relation $\text{IND}(A)$ can be certainly contained in X ; $A^-(X)$ is an object set whose objects belonging to the equivalence classes generated by the indiscernibility relation $\text{IND}(A)$ can be possibly contained in X . For any nonempty subset $X \subseteq U$ and $A \subseteq C$, X is a definable set of with respect to A iff $A_-(X) = A^-(X)$; otherwise, X is a rough set with respect to A iff $A_-(X) \neq A^-(X)$. The vagueness of X can be described by the accuracy of approximations:

$$\alpha_A(X) = \frac{|A_-(X)|}{|A^-(X)|}, \quad (5)$$

where $0 \leq \alpha_A(X) \leq 1$. $\alpha_A(X)$ provides a measure of how closely the lower and upper approximations of X .

The definition of accuracy of approximations can be applied to a partition or classification $F = \{X_1, X_2, \dots, X_n\}$ ($U = \bigcup_{i=1}^n X_i$) of U . Subsets, $X_i, i = 1, \dots, n$, are equivalence classes of U . The lower and upper approximations of F are expressed as $A_-(F) = \{A_-(X_1), A_-(X_2), \dots, A_-(X_n)\}$ and $A^-(F) = \{A^-(X_1), A^-(X_2), \dots, A^-(X_n)\}$, respectively. The accuracy of approximations of the classification F is defined as

$$\alpha_A(F) = \frac{|A_-(F)|}{|A^-(F)|} = \frac{\sum_{i=1}^n |A_-(X_i)|}{\sum_{i=1}^n |A^-(X_i)|}. \quad (6)$$

$\alpha_A(F)$ denotes the ratio of all correctly classified objects to possibly classified objects by means of attributes from A . It measures the classification ability with respect to A . The closer to 1 the value of $\alpha_A(F)$ is, the greater is the classify accuracy with respect to A .

2.3. Binary Ant Colony Optimization. Ant colony optimization (ACO), initially proposed by Dorigo et al., is a stochastic metaheuristic for solutions to global combinatorial optimization problems [27]. ACO is a nature-inspired intelligent algorithm inspired from the phenomenon that ants could always find the shortest path between the nest and food. The main idea of ACO is that a number of ants cooperate according to the pheromone laid on the path. ACO has been successfully applied in data mining, particularly for learning classification rules [28]. Binary ant colony optimization (BACO) is a binary version of ACO to solve the binary optimization problems. For BACO, the artificial ants walk on the mapping graph to represent the given binary optimization problem as described in Figure 1. And each

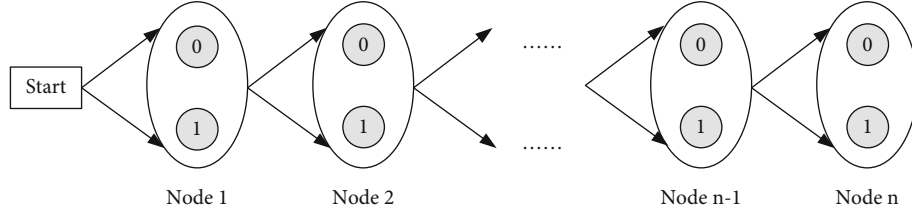


FIGURE 1: Path diagram of a binary ant colony optimization.

path corresponds to a potential solution to the given problem [29].

Ant selects the path $ij (j \in \{0, 1\})$ to walk from node 1 to node n by depending on the pheromone on the path. τ_{ij} represents the pheromone level for path $ij (j \in \{0, 1\})$. The solutions of BACO can be represented by a binary string: $x = \{x_1, x_2, \dots, x_n\}, x_i \in \{0, 1\}$. If the ant selects path $i0$, then $x_i = 0$, and if the ant selects path $i1$, then $x_i = 1$. The selection probability of path $ij (j \in \{0, 1\})$ for each ant can be defined as [30]

$$p_{i0} = \frac{\tau_{i0}^\alpha * \eta_{i0}^\beta}{\tau_{i0}^\alpha * \eta_{i0}^\beta + \tau_{i1}^\alpha * \eta_{i1}^\beta}, \quad (7)$$

$$p_{i1} = 1 - p_{i0}, \quad (8)$$

where η_{i0} and η_{i1} represent the visibility of path $i0$ and $i1$, and α and β are both weight factors.

Every ant makes a complete selection of the entire paths to generate a solution. After all ants complete the routines, all the solutions generated during the current iteration are evaluated by the fitness function $f(x)$. The ants having better fitness values are selected to enhance the pheromone. The pheromones on each path are updated as

$$\tau_{ij} = (1 - \rho)\tau_{ij} + \sum_{k=1}^T \Delta\tau_{ij}^k, \quad (9)$$

where ρ is the evaporation rate of pheromone, and $\Delta\tau_{ij}^k$ is the intensified pheromone of ant k on path ij .

3. BACORS

Optimal discretization of multiple attributes is an actual subset optimization problem, which selects optimal cut point set from the candidate cut point set. The solution of the problem can be represented by a binary-encoding bit string. For each candidate cut point, if it belongs to the optimal cut point set, the bit value is 1; otherwise, the bit value is 0. Obviously, it is a binary optimization problem. As BACO is a useful technique to solve binary optimization problem, it is used to deal the optimal discretization problem of multiple attributes in this article. The selection of optimal cut point set with BACO during the discretization process can also be described in Figure 1, where candidate cut points are represented by the node 1 to N . The length of the routine is equal to the cardinal number of the candidate cut point

set. If ant walks through the upper path $i0$, it means that the candidate cut point i is not selected. And if the lower path $i1$ is selected by the ant, the candidate cut points i is selected. Pheromone laid on the path is used for selections of 0 and 1 for each bit of the solution string. And the probabilities of the selections of 0 and 1 are calculated depending on the pheromone.

The fitness function $f(x)$ is used to evaluate the solution (ant) x in BACO. To find the optimal cut point set, the fitness function of the solution is defined as

$$f(x) = Q \cdot \frac{\alpha_B(F_x)}{|C_x|^N}, \quad (10)$$

where $Q (Q > 0)$ and $N (N > 0)$ are both constants, $|C_x|$ is the cardinal number of cut point set C_x , and $\alpha_B(F_x)$ is the accuracy of approximation of classification F_x induced by C_x . For multiple attributes, the optimal discretization is to find the least cut points among all attributes by keeping consistency for class. Therefore, the construction of fitness function takes into account the consistency and the number of cut points. The consistency is measured by $\alpha_B(F_x)$ defined in rough set. The solution of the ant corresponding to higher accuracy of approximation and less cut points has higher fitness value.

An ant deposits various amounts of pheromone on the path, which can be detected by other ants. The path having richer pheromone concentration will be selected by more ants. The laid pheromone of an ant k on path ij is defined as

$$\Delta\tau_{ij}^k = \begin{cases} f(x) & \text{if ant } k \text{ walks through } ij \\ 0 & \text{otherwise} \end{cases}. \quad (11)$$

The pheromone on each path are updated for each ant i . And the selection probability of paths $i0$ and $i1$ for each ant is calculated by using Eq. (12). To improve the search ability, some randomness is added to calculate the transfer direction of the ant. The value of the solution is determined as

$$x_i = \begin{cases} 0 & p_{i0} \cdot r_0 < p_{i1} \cdot r_1 \\ 1 & \text{else} \end{cases}. \quad (12)$$

where r_0 and r_1 are both random numbers in $[0, 1]$. The steps of BACORS are as follows:

Input: information system $S = \langle U, CUD, V, f \rangle$;
Output: cut point set C_{opt} corresponding to the solutions having highest fitness values;
Initialize parameters of BACORS including the population of ant colony P , the number of generations G , best fitness f ; pheromone τ_0 of each ant on each path, visibility of the path η , evaporation rate of pheromone ρ and accuracy of approximation classification μ defined in rough set;
Calculate the candidate cut point set $C^{candidate}$ of multiple continuous attributes C ;
Determine the solution of each ant $x = \{x_1, x_2, \dots, x_{|C^{candidate}|}\}, x_i \in \{0, 1\}$;
for $m=1, 2, \dots, P$
 for $n=1, 2, \dots, G$
 Determine the routine of ant n by Eq. (7) and Eq. (8);
 Obtain the cut point set C_x according to the routine;
 Calculate the accuracy of approximation classification $\alpha(F_x)$ by Eq. (6);
 Calculate the fitness $f(n)$ of ant n by Eq. (10);
 if $f(n) > f$ then
 Determine the solution of ant n $x^{(n)} = \{x_1, x_2, \dots, x_{|C^{candidate}|}\}, x_i \in \{0, 1\}$;
 Obtain the cut point set C_{opt} corresponding to the solution;
 end if
 if $\alpha(F_x) > \mu$ then
 Calculate the intensified pheromone $\Delta\tau$ of ant n by Eq. (11);
 end if
 end for
 Updating the pheromone on the paths by Eq. (9);
end for

ALGORITHM 1

4. Experimental Study

4.1. Data Sets. A wireless indoor localization data set is gathered from the U.C. Irvine repository. It is collected for the purpose of performing experimentation on how signal strengths can be used to determine one of the indoor locations of the users. The data set totally have 2000 instances and eight attributes. The first seven attributes correspond to the signal strengths captured by Android devices from seven routers. And the last attribute is the class label including four numbers 1, 2, 3, and 4 which correspond to the four locations of the office such as conference room, kitchen, or the indoor sports room.

4.2. Experiment Settings. The performance of our proposed method is compared with seven other commonly used discretization methods:

- (1) EW (equal width)
- (2) EF (equal frequency)
- (3) MDLP (minimal description length principle)
- (4) ChiMerge
- (5) Chi2
- (6) CAIM (class-attribute interdependency maximization)
- (7) CACC (class-attribute contingency coefficient)
- (8) BACORS (hybrid BACO and RS)

Among the eight discretization methods, EW and EF are unsupervised and require the user to specify the number of

intervals. It is set to 5 and 10 in our experiment. The last six methods are supervised discretization methods. MDLP is an entropy-based discretization method. ChiMerge and Chi2 are χ^2 measure-based discretization methods. For the ChiMerge method, we set the level of significance to 0.95. CAIM and CACC are dependency-based methods.

To validate the effectiveness of different discretization methods, the C4.5, Naive Bayes, and Bayes Network classifiers are implemented to the experimental data set. It has been proved that discretization leads to the improvement of accuracy and efficiency of the three classifiers. This is the reason why the three classifiers were chosen. The wireless indoor localization data set is classified into training sets and test sets. The instances of training sets are selected randomly from the original data, and five training and testing schemes are generated. The ratios of original data for generating five trainings are set to 0.1, 0.15, 0.2, 0.25, and 0.3, respectively. The discretization methods are applied to the training sets, and the testing sets are discretized using the generated cut points from the training set. The classification accuracies as a major evaluation indicator are compared among the eight discretization methods.

4.3. Results. The wireless indoor localization data are discretized with the proposed BACORS and other different discretization methods. The impact of the discretization methods on the classification result will be analyzed. Here, the original and discretized wireless indoor location data are classified by C4.5, Naive Bayes, and Bayes Network classifiers. The classification accuracies of the three classifiers with different discretization methods are shown in Tables 1–3. To effectively reflect the differences of the classification accuracies in Tables 1–3, four maps are drawn (Figures 2–5), which can give direct

TABLE 1: Classification accuracies of C4.5 classifier for original and discretized wireless indoor localization data with different discretization methods.

Methods	Ratios					Mean
	0.10	0.15	0.20	0.25	0.30	
Original	93.33	94.41	95.31	96.40	95.64	95.02
EW (10)	93.67	93.65	93.50	93.13	94.93	93.78
EW (5)	89.06	91.59	91.88	89.60	91.14	90.65
EF (10)	82.94	90.64	91.06	92.80	93.86	90.26
EF (5)	87.89	92.59	92.31	92.27	94.79	91.97
MDL	91.11	94.24	94.50	94.13	95.36	93.87
CAIM	91.70	94.47	94.50	94.40	96.14	94.24
CACC	92.22	91.29	94.50	94.40	96.21	93.72
ChiMerge	92.89	95.24	95.13	95.20	95.21	94.73
Chi2	93.17	94.71	92.56	96.20	95.64	94.46
BACORS	95.78	95.94	96.06	96.87	96.79	96.29

TABLE 2: Classification accuracies of Naive Bayes classifier for original and discretized wireless indoor localization data with different discretization methods.

Methods	Ratios					Mean
	0.10	0.15	0.20	0.25	0.30	
Original	94.72	96.82	96.94	97.27	96.86	96.52
EW (10)	93.83	94.47	96.75	96.47	96.64	95.63
EW (5)	90.67	90.24	92.75	93.13	93.86	92.13
EF (10)	93.00	93.00	95.56	95.87	96.36	94.76
EF (5)	92.06	94.59	94.31	95.60	95.21	94.35
MDL	95.72	96.47	96.06	96.80	96.78	96.37
CAIM	95.06	95.65	96.06	95.87	95.71	95.67
CACC	96.00	95.88	96.56	96.27	96.00	96.14
ChiMerge	94.56	96.12	96.75	97.40	96.50	96.27
Chi2	95.44	95.76	95.63	96.27	97.14	96.05
BACORS	96.78	97.06	97.38	97.53	97.50	97.25

TABLE 3: Classification accuracies of Bayes Network classifier for original and discretized wireless indoor localization data with different discretization methods.

Methods	Ratios					Mean
	0.10	0.15	0.20	0.25	0.30	
Original	94.72	95.80	96.88	96.27	96.79	96.09
EW (10)	95.56	95.94	96.63	96.65	97.07	96.37
EW (5)	91.13	91.00	91.63	92.13	92.86	91.75
EF (10)	95.11	95.35	96.38	96.27	97.57	96.14
EF (5)	94.94	95.06	94.81	94.87	95.79	95.09
MDL	95.44	95.76	96.69	96.60	96.79	96.26
CAIM	95.50	95.76	96.50	96.20	97.14	96.22
CACC	95.00	95.47	96.38	96.20	97.36	96.08
ChiMerge	95.78	96.35	96.56	97.07	97.93	96.74
Chi2	95.50	95.53	96.25	96.73	97.36	96.27
BACORS	96.39	96.76	97.31	97.27	97.64	97.07

and visual expressions for the changes of the classification accuracies with different discretization methods.

4.4. Discussions. In our experimental study, BACORS is used to discretize multiple continuous wireless indoor location data for improving the classification accuracies. To assess the effectiveness, the classification accuracies of C4.5, Naive Bayes, and Bayes Network classifiers are compared between BACORS and other commonly used discretization methods. Figure 2 displays a map of the mean classification accuracies for the original and discretized wireless indoor localization data with different discretization methods. The mean classification accuracies are calculated for the five training and testing schemes. It can be seen that discretization methods have obviously impacts on the performance of the classifiers. Compared to the original data, the classification results are improved when the data are discretized with some discretization methods, but the performance of the discretization methods is distinct when the discretized data are classified by different classifiers. The mean classification accuracies of C4.5 and Naive Bayes classifiers for the discretized data with other commonly used discretization methods are lower than the mean classification accuracy for original data. However, the classification accuracies of Bayes Network classifier are improved when EW (10), EF (10), MDL, CAIM, ChiMerge, and Chi2 are used to discretize the original data. The mean classification accuracies for the discretized data with BACORS are 96.29%, 97.07%, and 97.25% when the three classifiers are used. They are the highest among the classification accuracies for the original data or discretized data with other discretization methods.

Figures 3–5 display maps of the classification accuracies of C4.5, Naive Bayes, and Bayes Network, respectively, when different discretization methods are applied to the five types of training and testing schemes. From these three maps, the impacts of discretization methods on the classification results are compared. From Figure 3, it can be seen that the classification accuracies of C4.5 classifier are the highest when BACORS is applied to all the five training and testing scheme. And they increase when the sizes of training set become larger. Other discretization methods have bad performance especially for unsupervised discretization methods EW and EF. The classification accuracies of C4.5 classifier based on those discretization methods are lower than the classification accuracies for the original data. The number of interval is a parameter for EW and EF which needs to be predefined before discretization. As there is no rule to find the optimal number of intervals, 5 and 10 are set in our article. Compared to EW and EF, the classification accuracies of C4.5 classifier are significantly higher when BACORS is used.

Figure 4 is also analyzed and showed that the differences of classification accuracies become smaller for Naive Bayes classifier. The classification accuracies are larger than that for original data when the ratio for training set is 0.1, and MDL, CAIM, CACC, and Chi2 are used for discretization. Meanwhile, the classification accuracies are also improved when ChiMerge and Chi2 are used and the ratios for training set are 0.25 and 0.3, respectively. The classification

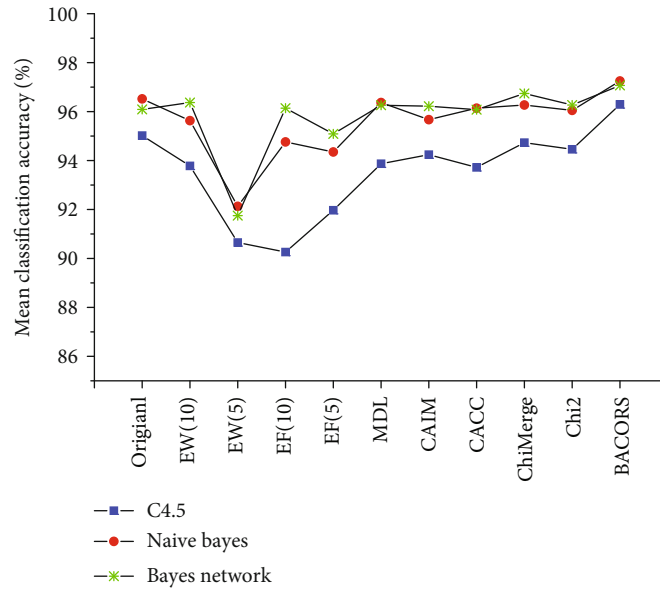


FIGURE 2: Mean classification accuracies of the three classifiers for original and discretized wireless indoor localization data with different discretization methods.

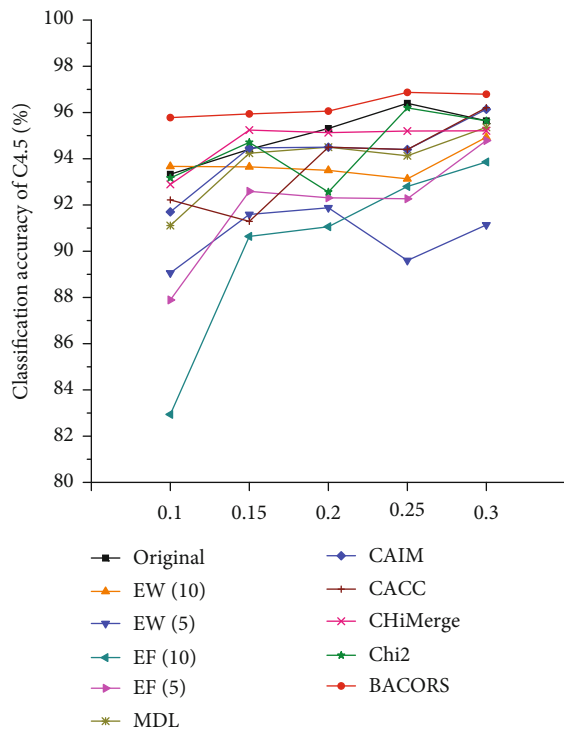


FIGURE 3: Classification accuracies of C4.5 classifier for original and discretized wireless indoor localization data with different discretization methods.

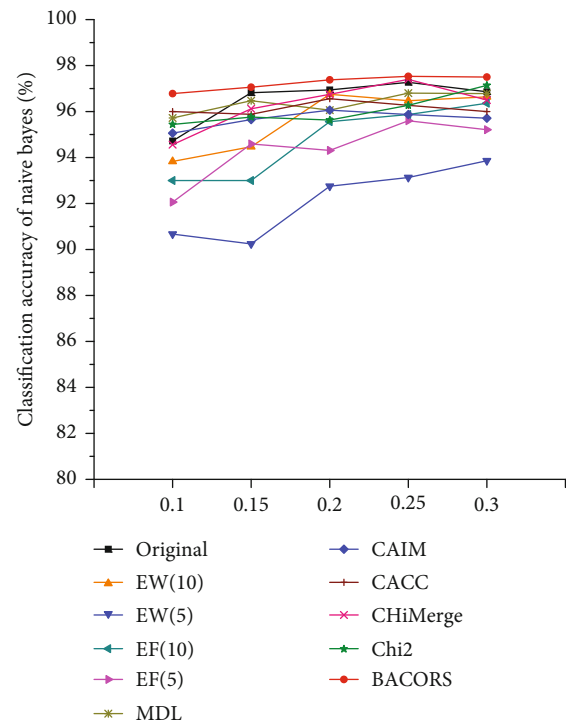


FIGURE 4: Classification accuracies of Naive Bayes classifier for original and discretized wireless indoor localization data with different discretization methods.

accuracies of Naive Bayes classifier are still the highest for all the training and testing schemes discretized based on BACORS. It can be seen that from Figure 5, the classification accuracies of Bayes Network are very close for most of the discretization methods. When the ratios for training set are 0.1, 0.25, and 0.3, the classification accuracies based on most

of the discretization methods are much larger than that for original data. It indicates that discretization can help to improve the efficacy of the classification result for Bayes Network classifier. For BACORS, the classification accuracies are the highest except that the ratio for training set is 0.3, for which the highest accuracy is reached for ChiMerge

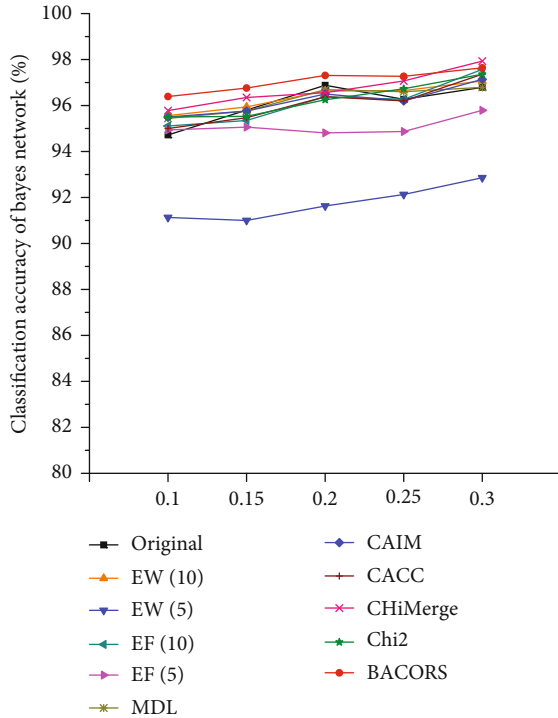


FIGURE 5: Classification accuracies of Bayes Network classifier for original and discretized wireless indoor localization data with different discretization methods.

and second highest for BACORS. Following the comparison of classification accuracies, we regard that BACORS is an optimal discretization method for enhancing the accuracy of indoor localization detection.

5. Conclusions and Future Works

In this paper, we propose a novel discretization method titled BACORS to construct the optimal granularity structures for multiple continuous attributes by combining the binary ant colony optimization with rough set. BACORS uses the binary ant colony optimization (BACO) algorithm to search the global optimal cut point set of multiple continuous attributes. Meanwhile, the accuracy of approximation classification defined in rough set is used to measure the consistency of the data and construct the fitness function to guide the path selection of the ant. We also apply BACORS to the wireless indoor localization data to validate the effectiveness. Comparing with several typical discretization methods, it can be seen that the BACORS has relative better performance on the classification of C4.5, Naive Bayes, and Bayes Network classifier. However, there are some issues to be resolved. For example, some parameters must be initialized while BACO is used for searching the optimal cut point set such as colony size, generation size, evaporation rate, and visibility of path. The efficiency of BACO depends on the selection of the parameters. How to select optimal BACO parameters to obtain better discretization results needs further research. For BACORS, the accuracy of approximation classification defined in rough set is

used to measure the consistency of multiattribute data. Multigranulation rough set (MGRS) is an extended version of classic rough sets, which uses multiple binary relations to construct granular structures rather than single binary relation. It is a kind of new information fusion strategies and has become one of desirable direction in granular computing [31–33]. In future researches, we will investigate more discretization methods based on BACO and MGRS to extend the applicability of the proposed method [34, 35].

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grant Nos. 62072291, 62072294, 61672332, 62002210), Anhui Provincial Natural Science Foundation (2008085MF202), Natural Science Foundation of Shanxi Province, China (Grant No. 20210302123455), and Open Project Program of the Key Laboratory of Embedded System and Service Computing Ministry of Education (Tongji University) (Grant No. 2021-04).

References

- [1] S. He and S. G. Chan, “Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.
- [2] J. G. Rohra, B. Perumal, S. J. Narayanan, P. Thakur, and R. B. Bhatt, “User localization in an indoor environment using fuzzy hybrid of particle swarm optimization & gravitational search algorithm with neural networks,” in *Proceedings of Sixth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing*, vol. 546, pp. 286–295, Singapore, 2017.
- [3] S. K. Jung, B. C. Moon, and S. Han, “Unsupervised learning for crowdsourced indoor localization in wireless networks,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2892–2906, 2016.
- [4] J. H. Bai, Y. L. Sun, W. X. Meng, C. L. Li, and X. Liu, “Wi-Fi fingerprint-based Indoor mobile user localization using deep learning,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6660990, 12 pages, 2021.
- [5] L. Pei, J. Liu, R. Guinness, Y. Chen, H. Kuusniemi, and R. Chen, “Using LS-SVM based motion recognition for smartphone indoor wireless positioning,” *Sensors*, vol. 12, no. 5, pp. 6155–6175, 2012.
- [6] X. Wang, L. Gao, and S. Mao, “PhaseFi: phase fingerprinting for indoor localization with a deep learning approach,” in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, San Diego, CA, USA, December 2015.
- [7] W. Zhang, K. Liu, W. D. Zhang, Y. M. Zhang, and J. Gu, “Deep neural networks for wireless localization in indoor and

- outdoor environments,” *Neurocomputing*, vol. 194, pp. 279–287, 2016.
- [8] M. Zhang, W. B. Shen, Z. F. Yao, and J. H. Zhu, “Multiple information fusion indoor location algorithm based on WIFI and improved PDR,” in *2016 35th Chinese Control Conference (CCC)*, pp. 5086–5092, Chengdu, China, July 2016.
- [9] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, and S. J. Kwon, “Outlier detection in indoor localization and internet of things (IoT) using machine learning,” *Journal of Communications and Networks*, vol. 22, no. 3, pp. 236–243, 2020.
- [10] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, “Granular computing: perspectives and challenges,” *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977–1989, 2013.
- [11] Z. Q. Yang, “An analysis of the application of granular computing in private data protection,” in *Knowledge Discovery and Data Mining*, pp. 497–502, Springer, Berlin, Heidelberg, 2012.
- [12] M. M. Eissa, M. Elmogy, and M. Hashem, “Rough - granular computing knowledge discovery models for medical classification,” *Egyptian Informatics Journal*, vol. 17, no. 3, pp. 265–272, 2016.
- [13] H. B. Liu, X. Y. Dao, and H. P. Guo, “Quantitative analysis for image segmentation by granular computing clustering from the view of set,” *Journal of Algorithm & Computational Technology*, vol. 13, 2019.
- [14] L. F. Ren and W. J. Wang, “A granular SVM-based method for top-N web services recommendation,” *IEEE Transactions on Services Computing*, vol. 15, no. 1, pp. 457–469, 2022.
- [15] F. Cao, Y. Ge, and J. F. Wang, “Optimal discretization for geographical detectors-based risk assessment,” *GIScience and Remote Sensing*, vol. 50, no. 1, pp. 78–92, 2013.
- [16] F. Cao, Y. Ge, and J. F. Wang, “Spatial data discretization methods for geocomputation,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, no. 10, pp. 432–440, 2014.
- [17] H. Liu, F. Hussain, C. L. Tan, and M. Dash, “Discretization: a enabling technique,” *Data Mining and Knowledge Discovery*, vol. 6, no. 4, pp. 393–423, 2002.
- [18] U. M. Fayyad and K. B. Irani, “On the handling of continuous-valued attributes in decision tree generation,” *Machine Learning*, vol. 8, no. 1, pp. 87–102, 1992.
- [19] J. Dougherty, R. Kohavi, and M. Sahami, “Supervised and unsupervised discretization of continuous features,” in *Machine learning proceedings*, pp. 194–202, California, July 1995.
- [20] Y. Ge, F. Cao, and R. F. Duan, “Impact of discretization methods on the rough set-based classification of remotely sensed images,” *International Journal of Digital Earth*, vol. 4, no. 4, pp. 330–346, 2011.
- [21] R. Kerber, “Chimerge: discretization of numeric attributes,” in *Proceeding of the tenth National Conference Artificial Intelligence on Artificial Intelligence*, pp. 123–128, AAAI Press, California, 1992.
- [22] H. Liu and R. Setiono, “Chi2: feature selection and discretization of numeric attributes,” in *Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence*, pp. 388–391, Herndon, VA, USA, November 1995.
- [23] L. A. Kurgan and K. J. Cios, “CAIM discretization algorithm,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2, pp. 145–153, 2004.
- [24] C. J. Tsai, C. I. Lee, and W. P. Yang, “A discretization algorithm based on class-attribute contingency coefficient,” *Information Sciences*, vol. 178, no. 3, pp. 714–731, 2008.
- [25] S. H. Nguyen and A. Skowron, “Quantization of real value attributes: rough set and boolean reasoning approach,” in *Proceedings of the Second Joint Annual Conference on Information Sciences, Wrightsville Beach, North Carolina*, pp. 34–37, Warsaw, Poland, October 1995.
- [26] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko, “Rough sets,” *Communications of the ACM*, vol. 38, no. 11, pp. 88–95, 1995.
- [27] M. Dorigo, M. Birattari, and T. Stützle, “Ant colony optimization,” *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2006.
- [28] J. Albinati, S. E. L. Oliveira, F. E. B. Otero, and G. L. Pappa, “An ant colony-based semi-supervised approach for learning classification rules,” *Swarm Intelligence*, vol. 9, no. 4, pp. 315–341, 2015.
- [29] A. Nadia, “Binary ant colony optimization for subset problems,” in *Multi-objective Swarm Intelligence*, pp. 105–121, Springer, Berlin Heidelberg, 2015.
- [30] W. Q. Xiong, L. Y. Wang, and C. Y. Yan, “Binary ant colony evolutionary algorithm,” *International Journal of Information Technology*, vol. 12, no. 3, pp. 10–20, 2006.
- [31] Y. H. Qian, J. Y. Liang, Y. Y. Yao, and C. Y. Dang, “MGRS: a multi-granulation rough set,” *Information Sciences*, vol. 180, no. 6, pp. 949–970, 2010.
- [32] C. Zhang, D. Li, and J. Liang, “Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes,” *Information Sciences*, vol. 507, pp. 665–683, 2020.
- [33] Y. H. Qian, H. Zhang, Y. L. Sang, and J. Y. Liang, “Multigranulation decision-theoretic rough sets,” *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 225–237, 2014.
- [34] C. Zhang, J. Ding, D. Li, and J. Zhan, “A novel multi-granularity three-way decision making approach in q-rung orthopair fuzzy information systems,” *International Journal of Approximate Reasoning*, vol. 138, pp. 161–187, 2021.
- [35] J. Zhan, H. Jiang, and Y. Y. Yao, “Three-way multiattribute decision-making based on outranking relations,” *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 10, pp. 2844–2858, 2021.