Applied Soft Computing Journal 89 (2020) 106087

Contents lists available at ScienceDirect

Applied Soft Computing Journal

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

Dynamic granularity selection based on local weighted accuracy and local likelihood ratio

Lei-Jun Li [a,b], Mei-Zheng Li [c,*], Ju-Sheng Mi [a,b], Bin Xie [c]

[a] College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, Hebei, 050024, PR China
[b] Hebei Key Laboratory of Computational Mathematics and Applications, Hebei Normal University, Shijiazhuang, Hebei, 050024, PR China
[c] College of Information Technology, Hebei Normal University, Shijiazhuang, Hebei, 050024, PR China

ARTICLE INFO

Article history:
Received 12 September 2018
Received in revised form 5 November 2019
Accepted 10 January 2020
Available online 16 January 2020

Keywords:
Granular computing
Dynamic granularity selection
Neighborhood granularity
Neighborhood rough set
Classification

ABSTRACT

Granular computing aims to develop a granular view for interpreting and solving problems, in which granularity selection is a key problem and has received extensive attention in recent years. Existing studies select the same granularity for all samples. In fact, different samples may prefer to different granularities. To address this issue, dynamic granularity selection is proposed in this paper. Namely, granularity selection is considered with respect to specific sample. Two indices, denoted as local weighted accuracy and local likelihood ratio, are introduced to compute the weight of granularity. Subsequently, an algorithm called DGS_LWA-LLS is given for dynamic granularity selection, in which the granularity with the largest weight is considered to be optimal The weight of granularity is related to specific sample, thus the weights of a granularity may be different with different samples. Consequently, different granularities will be selected with respect to different samples. Experiments were carried out based on neighborhood granularity to explain the necessity of granularity selection and to validate the rationality and effectiveness of DGS_LWA-LLS.

1. Introduction

Granular computing, proposed by Zadeh [1], focuses on knowledge representation and reasoning with information granules. It is employed to solve problems described with incomplete, uncertain, or vague information [2]. Three basic issues in granular computing are information granulation, organization and causation. As it was pointed out in [1,3,4], information granulation involves decomposition of whole into parts, organization involves integration of parts into whole, and causation involves association of causes with effects. As one of human advanced cognitive mechanisms, granular computing has become a fast growing field of research [5–9].

Granulation of an object set leads to a collection of granules. The size of granule is one of the essential ingredients for developing a theory of granular computing [10–12]. In 1985, Hobbs [13] introduced a concept of granularity to measure the size of granule. Granularity characterizes the granulation degree of objects from the viewpoint of hierarchy [1,14]. It represents discernibility ability of information in a granular structure [15]. The smaller the information granularity, the stronger its discernibility ability [16]. Pedrycz et al. [11] introduced a principle of justifiable granularity which delivered a conceptual and algorithmic vehicle to design an information granule on a basis of some experimental evidence.

How to select a proper granularity plays a crucial role for the utilization of granules in problem solving [17,18]. It has attracted increasing interest in recent years. To date, granularity selection has been studied according to various views and targets. Wu et al. [19] introduced multi-scale information table with different granulation levels. Subsequently, the issue of optimal scale selection has been investigated with different requirements [20–23]. In multi-scale information table, each object under each attribute is represented by different scales at different levels of granulations whose granular information transforms from a finer to a coarser labeled value. Zhu et al. [24] focused on neighborhood granularity and proposed a technique to select neighborhood granularity in classification learning based on the optimization of margin distribution. Zhao et al. proposed an optimal cost-sensitive granularization based on rough sets for variable costs [25] and designed adaptive neighborhood granularity based on the cost-sensitive feature selection with multi-level confidence [26]. Liu et al. [27] explored the granularity selection for cross-validation (CV) of support vector machine (SVM) in which CV is used for selecting the hyper-parameters of SVM. In [27], each fold of data set of CV can be considered as an information granule, and granularity selection is equal to the selection of the number of folds. Recently Xu et al. [28] studied

* Corresponding author.
E-mail addresses: lileijun1985@163.com (L.-J. Li), llimeizhengfirst@163.com (M.-Z. Li), mijsh@263.net (J.-S. Mi), xiebin_hebtu@126.com (B. Xie).

https://doi.org/10.1016/j.asoc.2020.106087
1568-4946/

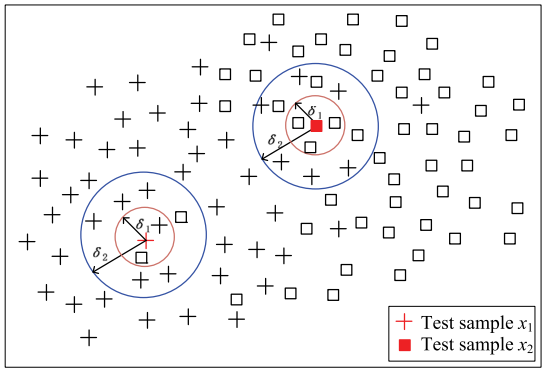

Fig. 1. An example of selecting different granularities for different samples.

how to select optimal granularity in generalized multigranulation rough sets.

However, existing studies select the same granularity for all samples. To the best of our knowledge, little research has considered granularity selection with respect to specific sample. In fact, different samples may prefer to different granularities. It is important to select a proper granularity in the classification of a given sample. Take neighborhood granularity as an example, it extracts information granules by computing the neighborhoods of samples. Different neighborhood granularities can be induced by different neighborhood sizes. Thus neighborhood size can be considered as a parameter to control the granularity of data analysis. As shown in Fig. 1, the test sample x_1 prefers granularity δ_2, while the test sample x_2 prefers granularity δ_1. Dynamic granularity selection is more targeted, thus it may be better than selecting the same granularity for all samples.

This paper aims to dynamically select an optimal granularity for each specific sample. Firstly, the weights of different granularities are computed for a given sample. Subsequently, the granularity with the largest weight is considered to be optimal for the sample. Note that the granularity weight is related to specific sample. Thus, the weights of a granularity may be different with different samples. Consequently, the optimal granularity will be different for different samples.

It is critical to compute the weight of granularity in the proposed method. As is known, the information granule is constructed based on the closeness among samples in terms of some useful spatial, temporal or functional relationships [3,29]. The samples are considered to be similar in an information granule. Thus, in this paper, the weight of granularity is computed based on information granule.

To be more specific, local weighted accuracy and local likelihood ratio are proposed based on information granule. The two indices are employed to evaluate the quality of granularity with respect to specific sample. The former evaluates the classification quality in information granule, while the latter reflects the locality of sample distribution in information granule.

The main contributions of this work include: (1) Dynamic granularity selection is proposed with respect to specific sample; (2) local weighted accuracy and local likelihood ratio are introduced. Subsequently, an algorithm called DGS_LWA-LLS is given for dynamic granularity selection; (3) a set of experiments are presented to validate the rationality and effectiveness of DGS_LWA-LLS. The great potential of dynamic granularity selection is also shown.

The rest of this paper is organized as follows. In Section 2, related work with granular computing, neighborhood granularity as well as neighborhood rough sets are recalled. In Section 3, local weighted accuracy and local likelihood ratio are introduced and a dynamic granularity selection method is proposed. Then we explain the necessity of granularity selection, validate the rationality and effectiveness of the proposed method, and present our experimental results in Section 4. Finally, conclusions and future work are given in Section 5.

2. Preliminaries

Granularity refers to the degree of granulation. Our aim is to select a proper granularity for a given sample in classification learning.

The granulation can be induced by a binary relation. Different binary relations lead to different granulation models. For example, an equivalent relation can induce a set of equivalence classes in which each equivalence class is regarded as an equivalence granule [7,30]; a tolerance relation can induce a set of tolerance classes in which each tolerance class is referred to as a tolerance granule [7,31]. Of these, neighborhood granulation provides an effective granular computing model in dealing with heterogeneous data. It has been successfully applied to various fields [32–34]. In this paper, we take neighborhood granularity as an example to illustrate our granularity selection method. In what follows, some basic notions and properties are reviewed. For more details please refer to [35,36].

Formally, the samples are expressed as a decision information system $\langle U, C, D\rangle$ in classification learning, where U is a finite nonempty set of samples, C is a set of condition attributes describing the samples, and D is a decision attribute indicating the classes of the samples.

Definition 1 ([35]). Given $x_i \in U$ and $B \subseteq C$, the neighborhood $\delta_B(x_i)$ of x_i with respect to B is defined as $\delta_B(x_i) = \{x_j | x_j \in U, \Delta_B(x_i, x_j) \leq \delta\}$, where Δ_B is a metric function defined in feature spaces B, i.e., $\forall x_1, x_2, x_3 \in U$, it satisfies:
(1) $\Delta_B(x_1, x_2) \geq 0$, $\Delta_B(x_1, x_2) = 0$ if and only if $x_1 = x_2$;
(2) $\Delta_B(x_1, x_2) = \Delta_B(x_2, x_1)$;
(3) $\Delta_B(x_1, x_3) \leq \Delta_B(x_1, x_2) + \Delta_B(x_2, x_3)$.

$\delta_B(x_i)$ is the neighborhood granule centered with x_i. Its neighborhood granularity is determined by the size of δ. The samples in a neighborhood granule are close to each other, and it is difficult to distinguish them.

In this paper, Heterogeneous Euclidean-Overlap Metric function (HEOM) [35,37] is used as metric function for mixed numerical and categorical data:

$$\mathrm{HEOM}(x, y) = \sqrt{\sum_{i=1}^{m} d_{a_i}^2(x_{a_i}, y_{a_i})}, \tag{2.1}$$

where m is the number of attributes, x_{a_i} and y_{a_i} are the attribute value of x and y with respect to attribute a_i, $d_{a_i}(x_{a_i}, y_{a_i})$ is the distance between samples x and y with respect to attribute a_i, defined as

$$d_{a_i}(x, y) = \begin{cases} 1, & \text{if } x_{a_i} \text{ or } y_{a_i} \text{ is unknown,} \\ overlap_{a_i}(x_{a_i}, y_{a_i}), & \text{if } a_i \text{ is a nominal attribute,} \\ rn\_diff_{a_i}(x_{a_i}, y_{a_i}), & \text{if } a_i \text{ is a numerical attribute,} \end{cases} \tag{2.2}$$

where $overlap_{a_i}(x_{a_i}, y_{a_i}) = \begin{cases} 0, & \text{if } x_{a_i} = y_{a_i}, \\ 1, & \text{otherwise,} \end{cases}$ and $rn\_diff_{a_i}(x_{a_i}, y_{a_i}) = \frac{|x_{a_i} - y_{a_i}|}{max_{a_i} - min_{a_i}}$.

Given a metric space $\langle U, \Delta_B\rangle$, the family of neighborhood granules $\{\delta_B(x_i) | x_i \in U\}$ forms an elemental granule system that

covers the universe. A neighborhood relation N_B on the universe can be written as a relation matrix $(r_{ij})_{n\times n}$ [35], where

$$r_{ij} = \begin{cases} 1, & \text{if } \Delta_B(x_i, x_j) \le \delta, \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Neighborhood relations are a kind of similarity relations, which satisfy the properties of reflexivity and symmetry [35]. Neighborhood relations draw the samples together for similarity in terms of distances. In addition, a neighborhood relation will degenerate into an equivalent relation if $\delta = 0$.

The influence of granularity on classification learning mainly includes two cases:

(1) Multi-granularity based classification

It means that the classification process is directly related to the size of granularity, such as neighborhood classifier (NEC) [36]. Given a sample, NEC first finds the neighborhood of the sample and then assigns the majority class of the neighborhood as its classification result. The detailed classification process is shown as Algorithm 1.

Algorithm 1 Neighborhood classifiers (NEC)

Require: $\langle U, C, D\rangle$: training set
δ: size of neighborhood granularity
x: test sample
Ensure: class of x
1: Compute the distances between x and every training sample with a given metric function
2: Find the samples in $\delta(x)$
3: Find the class with the majority training samples in $\delta(x)$ and assign it as the class of x

The key to NEC is δ, which determines the size of neighborhood granularity. There will be no neighborhood sample included in the neighborhood if δ is too small; on the other hand, the neighborhood cannot reflect the local information of the sample if δ is too great [36].

Besides, the classification process of K-nearest neighbor classifier (KNN) [38] is also based on granularity which is determined by the size of K.

(2) Multi-granularity subspaces based classification

Different granularity subspaces can be induced with different neighborhood granularities [35,36,39,40]. Accordingly, different classification performances can be obtained in different granularity subspaces. The representative is attribute reduction based on neighborhood rough set.

Rough set is a kind of representative granular computing model [41]. It has been proven to be an effective tool for attribute reduction [42–44]. The classical rough set model is based on equivalent relation and can only be used to evaluate categorical features. Lin introduced neighborhood relation into rough set methodology, and a neighborhood rough set model was constructed [45,46]. Neighborhood rough set has been successfully applied in heterogeneous attribute reduction [35,36,47].

Definition 2 ([35]). Given U and a neighborhood relation N_B over U, $\langle U, N_B\rangle$ is called a neighborhood approximation space. For any $X \subseteq U$, two subsets of samples, called lower and upper approximations of X, are defined as

$$\begin{aligned} \underline{\mathbf{N}_B}X &= \{x_i | \delta_B(x_i) \subseteq X, x_i \in U\}, \\ \overline{\mathbf{N}_B}X &= \{x_i | \delta_B(x_i) \cap X \neq \emptyset, x_i \in U\}. \end{aligned} \tag{2.4}$$

The order pair $\langle \underline{\mathbf{N}_B}X, \overline{\mathbf{N}_B}X\rangle$ is called a rough set of X with respect to the neighborhood relation N_B.

Definition 3 ([35]). Given a decision information system $\langle U, C, D\rangle$, D divides U into M equivalence classes: $X_1, X_2, \ldots, X_M$, $B \subseteq C$ generates the neighborhood relation N_B on U. Then the lower and upper approximations of D with respect to attributes B are defined as

$$\begin{aligned} \underline{\mathbf{N}_B}D &= \bigcup_{j=1}^{M} \underline{\mathbf{N}_B}X_j, \\ \overline{\mathbf{N}_B}D &= \bigcup_{j=1}^{M} \overline{\mathbf{N}_B}X_j. \end{aligned} \tag{2.5}$$

Definition 4 ([35]). Given a decision information system $\langle U, C, D\rangle$, the dependency degree of $B \subseteq C$ with respect to D is defined as

$$\gamma_B(D) = Card(\underline{\mathbf{N}_B}D)/Card(U). \tag{2.6}$$

Dependency degree is the percentage of samples in the decision positive region. It measures the approximate power of a subset of condition attributes with respect to the decision attribute. The range of dependency degree is [0,1]. We say D depends on B in the degree of $\gamma_B(D)$. The decision system is called consistent if $\gamma_B(D) = 1$. The monotonicity of dependency degree is discussed with condition attributes.

Property 1 ([35]). *Given δ and a decision information system $\langle U, C, D\rangle$, if $B_1 \subseteq B_2 \subseteq C$, then $\gamma_{B_1}(D) \le \gamma_{B_2}(D)$.*

The aim of attribute reduction is to find a subset of condition attributes which keeps the approximation ability of the whole set of condition attributes.

Definition 5 ([35]). Given a decision information system $\langle U, C, D\rangle$, $B \subseteq C$, if

$$\begin{aligned} &\gamma_B(D) = \gamma_C(D), \\ &\forall b \in B, \gamma_{B-b}(D) < \gamma_B(D), \end{aligned} \tag{2.7}$$

then B is called an attribute reduct.

Hu et al. further discussed the monotonicity of neighborhood dependency degree with different granularities.

Property 2 ([35]). *Given a decision information system $\langle U, C, D\rangle$, if $\delta_1 \le \delta_2$, then $\gamma_{\delta_2}(D) \le \gamma_{\delta_1}(D)$.*

δ has an impact on neighborhood dependency degree. Thus, different granularities lead to different attribute reducts, and B is also called a δ neighborhood separable subspace. Zhu et al. demonstrated that neighborhood attribute reduction is sensitive to the granularity in classification learning [24]. Besides, some forward feature subset selection algorithms are constructed to compute attribute reduct based on neighborhood dependency degree [35,48,49].

Finally, some work related to the selection of neighborhood granularity are reviewed.

The selection of neighborhood granularity has attracted considerable attention. It is acknowledged that a proper neighborhood granularity is important to classification learning [24,35,36,39,40]. However, it is still an open problem to select a proper neighborhood granularity. Most of the literatures are up to empirical values [35,36,39,40]. Specifically, the empirical values are estimated based on a series of experiments which are conducted on some data sets. As we know, different data sets may prefer to different neighborhood granularity. Thus the empirical values may be not suitable for other data sets.

In [24], Zhu et al. delivered a preliminary study to select neighborhood granularity in classification learning and proposed

an algorithm called GSC_MD. In GSC_MD, the weights of different granularity are computed based on optimization of margin distribution. Subsequently, the granularity with the largest weight is selected as the optimal granularity. The weight of granularity is related to specific data set, thus GSC_MD can adaptively select a proper granularity based on data set. That is to say, different granularities will be selected for different data sets. Compared with empirical values, GSC_MD is more targeted. Consequently, a better classification performance is obtained [24].

In GSC_MD, different granularities will be selected for different data sets, but all the samples in a data set use the same neighborhood granularity. In fact, different samples may prefer to different granularities. Thus, dynamic granularity selection is proposed in this paper, and granularity is selected with respect to specific sample.

3. Dynamic granularity selection based on local weighted accuracy and local likelihood ratio

In this section, local weighted accuracy and local likelihood ratio are introduced to evaluate the quality of granularity with respect to specific sample. And then an algorithm called DGS_LWALLS is given for dynamic granularity selection.

3.1. Local weighted accuracy

Given a test sample x and its neighborhood granule $\delta_i(x)$, local weighted accuracy $LWA_i(x)$ is defined as:

$$LWA_i(x) = \frac{\sum_{x_j \in \delta_i(x)} (1 - \frac{d_{jx}}{\delta_i}) \Lambda_{ij}}{\sum_{x_j \in \delta_i(x)} (1 - \frac{d_{jx}}{\delta_i})}, \tag{3.1}$$

where $x_j (x_j \neq x)$ is the neighborhood sample in $\delta_i(x)$, y_j is the true label of x_j, $\hat{y}_j^i$ is the classification result of x_j with granularity δ_i, $0 \leq d_{jx} \leq \delta_i$ is the distance between x and x_j, and

$$\Lambda_{ij} = \begin{cases} 1, & \text{if } y_j = \hat{y}_j^i, \\ 0, & \text{if } y_j \neq \hat{y}_j^i. \end{cases} \tag{3.2}$$

The impact of neighborhood sample x_j on $LWA_i(x)$ is determined by the distance between x and x_j. The smaller the value of d_{jx}, the higher the similarity between x and x_j, thus the greater the impact on $LWA_i(x)$. We can see that $LWA_i(x) = 1$ if all samples in neighborhood granule $\delta_i(x)$ are correctly classified and $LWA_i(x) = 0$ if all neighborhood samples are misclassified. In addition, $LWA_i(x)$ is denoted as 0 if there is no neighborhood sample in $\delta_i(x)$.

For $LWA_i(x)$, it evaluates the classification quality in neighborhood granule $\delta_i(x)$. We tend to select a granularity with larger local weighted accuracy. However, there are still some shortcomings to evaluate the quality of granularity δ_i with respect to x simply by using $LWA_i(x)$. Some other factors should also be considered:

(1) Number of neighborhood samples. Given a test sample x and its two neighborhood granules $\delta_1(x)$ and $\delta_2(x)$. As shown in Fig. 2, $\delta_1(x) \subseteq \delta_2(x)$. There are 1 and 10 neighborhood samples in $\delta_1(x)$ and $\delta_2(x)$, respectively. If all the samples are correctly classified in $\delta_1(x)$ and $\delta_2(x)$, then $LWA_1(x) = 1$ and $LWA_2(x) = 1$. In this case, there is no distinction between δ_1 and δ_2 based on local weighted accuracy. Consequently, the optimal granularity will be selected randomly from δ_1 and δ_2. However, more samples are contained in $\delta_2(x)$. It means that $\delta_2(x)$ is more stable than $\delta_1(x)$. In other words, δ_2 is more favorable than δ_1 with respect to x. It is coincide with the researches in Refs. [22,50], which tend to select a larger granularity.

(2) Number of classes in neighborhood granule. The samples are considered to be similar with each other in a neighborhood granule, thus their class labels tend to be similar. In fact, more consistent classes tend to mean a higher classification confidence. For example, neighborhood classifier assigns the class with the majority training samples in neighborhood granule as the classification result of test sample [36]. As shown in Fig. 3, the numbers of neighborhood samples in $\delta(x_1)$ and $\delta(x_2)$ are the same, but the number of classes in the two neighborhoods granules are different. Without loss of generality, suppose the number of classes in $\delta(x_1)$ and $\delta(x_2)$ is 1 and 2, respectively. In this case, it is considered that the classification result of x_1 has a higher confidence than that of x_2 based on neighborhood classifier.

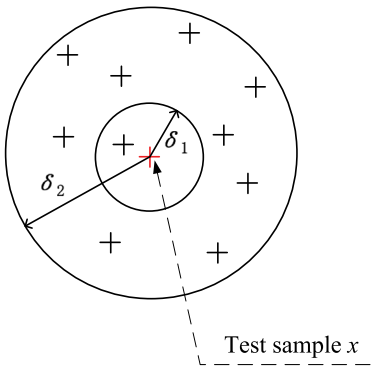

Fig. 2. An example of different neighborhood samples numbers.

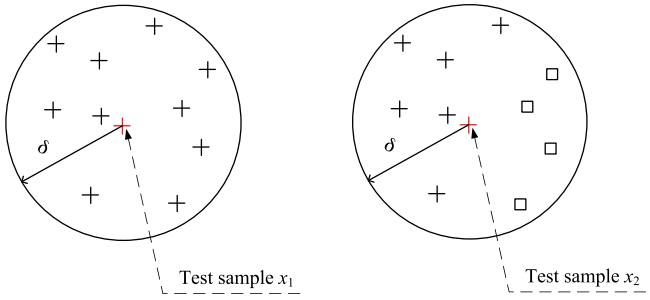

Fig. 3. An example of different numbers of classes in neighborhood granule.

(3) Types of neighborhood samples. As shown in Fig. 4, there are 100 training samples, of which 20 samples are with class "+1"(marked as "+ ") and the other are with class "−1"(marked as "□"). Given a test sample x and its neighborhood granule containing 10 samples, of which 5 samples are with class "+1" and 5 samples are with class "−1". Then how to determine the classification result of x based on the neighborhood samples? In other words, which type of neighborhood samples ("+ " or "□") has a greater impact on the neighborhood granule? As we know, if we select 10 samples at random, there should be 2 samples with class "+1" and 8 samples with class "−1". However, there are 5 samples with class "+1" and 5 samples with class "−1" in the neighborhood granule. The reason is that a neighborhood granule aims to reflect the locality of sample distribution. Thus, it is deemed that the samples with class "−1" has a greater impact than the samples with class "+1" in the neighborhood granule. Consequently, the classification result of x should be "−1" based on the neighborhood samples.

In what follows, local likelihood ratio is proposed to describe the above three factors.

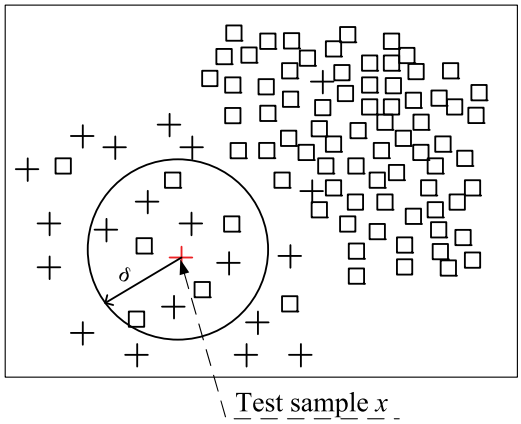

Fig. 4. An example of different types of neighborhood samples.

3.2. Local likelihood ratio

Given a test sample x and its neighborhood granule $\delta_i(x)$, local likelihood ratio $LLR_i(x)$ is defined as:

$$LLR_i(x) = \sum_{k=1}^{M} f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)}, \tag{3.3}$$

where M is the number of classes, $f_i^k(x)$ is the number of neighborhood samples with class label k in $\delta_i(x)$, $e_i^k(x)$ is the expected frequency of samples with class label k in $\delta_i(x)$ and is computed as:

$$e_i^k(x) = N_i(x) \times pr_k, \tag{3.4}$$

where $N_i(x)$ is the number of neighborhood samples in $\delta_i(x)$, and $0 < pr_k < 1$ is the proportion of samples with class label k in training set. In addition, if $f_i^k(x) = 0$, then $f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)} = 0$. Further, if $N_i(x) = 0$, then $LLR_i(x) = 0$. Namely, if a neighborhood is too small and there is no neighborhood sample in $\delta_i(x)$, then its local likelihood ratio is 0.

The global sample distribution is denoted as $Pr = [pr_1, \ldots, pr_M]$ in which $\sum_{k=1}^{M} pr_k = 1$. The local sample distribution with respect to neighborhood granule $\delta_i(x)$ is denoted as $Pr^{\delta_i(x)} = [pr_1^{\delta_i(x)}, \ldots, pr_M^{\delta_i(x)}]$ in which $0 \le pr_k^{\delta_i(x)} \le 1$ is the proportion of samples with class label k in $\delta_i(x)$ and $\sum_{k=1}^{M} pr_k^{\delta_i(x)} = 1$. According to Eq. (3.3), the following properties can be obtained.

Property 3. *Given a granularity δ_i and a test sample x, if $Pr = Pr^{\delta_i(x)}$, then $LLR_i(x) = 0$.*

Proof. $\forall k \in M$, if $f_i^k(x) = 0$, then $f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)} = 0$; otherwise, since $Pr = Pr^{\delta_i(x)}$, then $pr_k = pr_k^{\delta_i(x)}$ and $f_i^k(x) = N_i(x) \times pr_k^{\delta_i(x)} = N_i(x) \times pr_k = e_i^k(x)$. Consequently, $f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)} = f_i^k(x) \times \log_2 1 = 0$. Then we can conclude that $LLR_i(x) = \sum_{k=1}^{M} f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)} = 0$. □

Especially, if the neighborhood granule $\delta_i(x)$ includes all the training samples, then $LLR_i(x) = 0$. In fact, if a neighborhood is too great and all the training samples are included in $\delta_i(x)$, then there is no local information in the neighborhood granule. Thus its local likelihood ratio is 0.

Property 4. *Given two neighborhood granules $\delta_1(x)$ and $\delta_2(x)$ with respect to test sample x, if $Pr^{\delta_1(x)} = Pr^{\delta_2(x)}$ and $\delta_1(x) \subseteq \delta_2(x)$, then $LLR_1(x) \le LLR_2(x)$.*

Proof. $\delta_1(x) \subseteq \delta_2(x)$ implies $N_2(x) = p \times N_1(x)$ and $p = \frac{N_2(x)}{N_1(x)} \ge 1$. Further, if $Pr^{\delta_1(x)} = Pr^{\delta_2(x)}$, then $\forall k \in M$, $f_2^k(x) = N_2(x) \times pr_k^{\delta_2(x)} = p \times N_1(x) \times pr_k^{\delta_1(x)} = p \times f_1^k(x)$. Thus, $LLR_2(x) = \sum_{k=1}^{M} f_2^k(x) \log_2 \frac{f_2^k(x)}{e_2^k(x)} = \sum_{k=1}^{M} f_2^k(x) \log_2 \frac{N_2(x) \times pr_k^{\delta_2(x)}}{N_2(x) \times pr_k} = \sum_{k=1}^{M} p \times f_1^k(x) \log_2 \frac{N_1(x) \times pr_k^{\delta_1(x)}}{N_1(x) \times pr_k} = \sum_{k=1}^{M} p \times f_1^k(x) \log_2 \frac{f_1^k(x)}{e_1^k(x)} = p \times \sum_{k=1}^{M} f_1^k(x) \log_2 \frac{f_1^k(x)}{e_1^k(x)} = p \times LLR_1(x)$.

In what follows, we prove $LLR_1(x) \ge 0$.

$LLR_1(x) = \sum_{k=1}^{M} f_1^k(x) \log_2 \frac{f_1^k(x)}{e_1^k(x)}$. Without loss of generality, suppose $M = 2$. Namely, the number of classes is two.

(1) $N_1(x) = 0$. In this case $LLR_1(x) = 0$.

(2) $N_1(x) > 0$. In this case $f_1^1(x) + f_1^2(x) = e_1^1(x) + e_1^2(x) = N_1(x)$ and $0 \le f_1^1(x), f_1^2(x), e_1^1(x), e_1^2(x) \le N_1(x)$. The aim is to prove $LLR_1(x) = \sum_{k=1}^{2} f_1^k(x) \log_2 \frac{f_1^k(x)}{e_1^k(x)} = f_1^1(x) \log_2 \frac{f_1^1(x)}{e_1^1(x)} + (N_1(x) - f_1^1(x)) \log_2 \frac{N_1(x) - f_1^1(x)}{N_1(x) - e_1^1(x)} \ge 0$.

For convenience of discussion, denoted $f_1^1(x), e_1^1(x), N_1(x)$ as q, r, s, respectively.

$q \log_2 \frac{q}{r} + (s-q) \log_2 \frac{s-q}{s-r} \ge 0 \Leftrightarrow \log_2 (\frac{q}{r})^q (\frac{s-q}{s-r})^{s-q} \ge 0 \Leftrightarrow (\frac{q}{r})^q (\frac{s-q}{s-r})^{s-q} \ge 1 \Leftrightarrow q^q (s-q)^{s-q} \ge r^q (s-r)^{s-q} (0 \le q, r \le s)$.

Denote $h(z) = z^q (s-z)^{s-q}$. $h'(z) = s z^{q-1} (s-z)^{s-q-1} (q-z)$. It means that $h'(z) \ge 0$ when $0 \le z < q$, $h'(z) = 0$ when $z = q$, and $h'(z) \le 0$ when $q < z \le s$. Thus, $q^q (s-q)^{s-q} \ge r^q (s-r)^{s-q} (0 \le q, r \le s)$.

Based on the above analysis, it can be concluded that $LLR_1(x) \ge 0$. Further, $p \ge 1$, thus $LLR_2(x) = p \times LLR_1(x) \ge LLR_1(x)$. □

Property 4 shows the monotonicity of local likelihood ratio. It can be seen that large granularity tends to be selected when the same local sample distribution is given. In fact, local sample distribution is related to the number of classes and the types of neighborhood samples in a neighborhood granule. Specifically, given local sample distribution $Pr^{\delta_i(x)} = [pr_1^{\delta_i(x)}, \ldots, pr_M^{\delta_i(x)}]$ with respect to neighborhood granule $\delta_i(x)$ in which $0 \le pr_k^{\delta_i(x)} \le 1$, the number of classes in $\delta_i(x)$ determines the number of nonzero values, and the types of neighborhood samples determines the location of nonzero values in $Pr^{\delta_i(x)}$. In a nutshell, $LLR_i(x)$ takes into account the number of neighborhood samples in $\delta_i(x)$, the number of classes in $\delta_i(x)$ and the types of neighborhood samples. $LLR_i(x)$ is inspired by the likelihood ratio [51]. It provides an information-theoretic measure of the (noncommutative) distance of sample distribution between local neighborhood granule $\delta_i(x)$ and global training samples.

Neighborhood granule aims to reflect the locality of sample distribution. The greater the local likelihood ratio, the better the locality of the neighborhood granule. Therefore, we tend to select a granularity with larger local likelihood ratio.

Finally, a dynamic granularity selection method is given based on local weighted accuracy and local likelihood ratio.

3.3. DGS_LWA-LLS

Given a decision information system $\langle U, C, D \rangle$ with m different granularities $\delta_i (i = 1, 2, \ldots, m)$ and a test sample x. The aim is to select the optimal granularity with respect to x in classification learning. Fig. 5 shows the flow chart of the proposed method.

In Fig. 5, $w_i(x)$ is the weight of granularity δ_i with respect to x. It is employed to evaluates the quality of granularity δ_i with respect to x and is the key to the proposed method. As is known, neighborhood relations draw the objects together for similarity in terms of distances. The samples are considered to be similar with x in neighborhood granule $\delta_i(x)$. Thus, in this paper, the weight $w_i(x)$ is computed based on $\delta_i(x)$. To be more specific, the weight

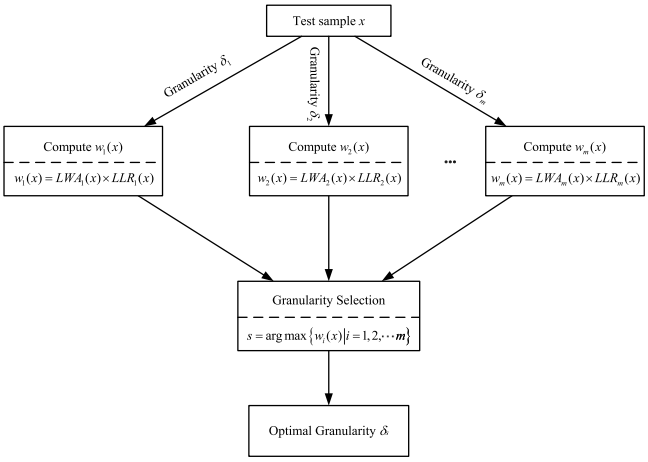

Fig. 5. Flow chart of the proposed method.

$w_i(x)$ is computed based on local weighted accuracy $LWA_i(x)$ and local likelihood ratio $LLR_i(x)$:

$$w_i(x) = LWA_i(x) \times LLR_i(x), \tag{3.5}$$

where local weighted accuracy can be deemed as a kind of empirical risk in neighborhood granule. A large local weighted accuracy implies a small empirical risk. As to local likelihood ratio, it reflects the locality of sample distribution in neighborhood granule.

After computing the weights, the granularity with the largest weight is considered to be optimal. Since $w_i(x)$ is based on a given sample x, it may be different with different samples. Consequently, the optimal granularity selected for different samples may be different. Since local weighted accuracy and local likelihood ratio are considered, the proposed dynamic granularity selection method is denoted as DGS_LWA-LLS. The pseudocodes of DGS_LWA-LLS is given as Algorithm 2.

Algorithm 2 DGS_LWA-LLS.

Require: $\langle U, C, D\rangle$: decision information system
$\delta_i (i = 1, 2, \ldots, m)$: m different granularities
x: test sample
Ensure: δ_s: the selected granularity with respect to x
1: Compute the distance d_{jx} between x and every training sample x_j in U
2: **for** $i = 1, 2, \ldots, m$ **do**
3: Find the neighborhood of x in granularity δ_i and denoted as $\delta_i(x) = \{x_j | x_j \in U, d_{jx} \le \delta_i\}$
4: Compute $LWA_i(x) = \frac{\sum_{x_j \in \delta_i(x)} (1 - \frac{d_{jx}}{\delta_i}) \Lambda_{ij}}{\sum_{x_j \in \delta_i(x)} (1 - \frac{d_{jx}}{\delta_i})}$ based on $\delta_i(x)$ and Eq. (3.1)
5: Compute $LLR_i(x) = \sum_{k=1}^{M} f_i^k(x) \log_2 \frac{f_i^k(x)}{e_i^k(x)}$ based on $\delta_i(x)$ and Eq. (3.3)
6: Compute $w_i(x) = LWA_i(x) \times LLR_i(x)$
7: **end for**
8: $s = \arg\max\{w_i(x) | i = 1, 2, \ldots, m\}$
9: Output δ_s

Here neighborhood granularity is used as an example to illustrate DGS_LWA-LLS. In fact, DGS_LWA-LLS not only can be applied to the selection of neighborhood granularity, but also be suitable for other granular computing models, such as equivalence information granularity [30], in which local weighted accuracy and local likelihood ratio are computed based on equivalence information granule.

Table 1
Statistics of classification tasks.

No.	Data set	Sample	Attribute	Class
1	breast	84	9216	5
2	crx	690	15	2
3	heart	270	13	2
4	hepatitis	155	19	2
5	iono	351	34	2
6	rice	104	5	2
7	sonar	208	60	2
8	spam	4601	57	2
9	wdbc	569	30	2
10	wine	178	13	3

4. Experimental evaluation and analysis

In this section, some experiments are conducted to explain the necessity of granularity selection and to validate the rationality and effectiveness of DGS_LWA-LLS. Table 1 summarizes 10 UCI (University of California at Irvine) data sets [52] used in this work. The numbers of samples range from 84 to 4601 and the numbers of attributes range from 5 to 9216. For every classification task, standard 10-fold cross validation is performed.

As mentioned in Section 2, the influence of granularity on classification learning mainly includes two cases: (1) multi-granularity based classification; (2) multi-granularity subspaces based classification. Thus the experiments are conducted from the two perspectives.

4.1. Multi-granularity based classification

Multi-granularity based classification means that the classification process is directly related to the size of granularity. Here we use NEC (Algorithm 1) as an example to test DGS_LWA-LLS.

First, we explain the necessity of granularity selection. In the experiment, we test the classification performances of NEC with different granularities varying from 0.01 to 1 with the step 0.01. As shown in Fig. 6, the classification accuracy of NEC varies greatly with granularity. The optimal classification performance occurs in different granularity. Neither too great nor too small granularity will lead to good classification performance. Thus it is necessary to choose a proper granularity to improve the classification performance.

Then, the rationality of DGS_LWA-LLS is analyzed. The weight of granularity is computed in DGS_LWA-LLS. The granularity with the largest weight is considered to be optimal. A question follows immediately: does larger granularity weight mean better prediction performance? The relationship between classification accuracy and weight of granularity is shown in Fig. 7. Here it should be noted that the weight of granularity is related to specific sample. Thus the weights of a granularity may be different with different samples. On the x-axis, 1 means the classification of every sample is based on the minimum weight and 100 means the classification of every sample is based on the maximal weight. Fig. 7 shows the trend that larger weight usually means better prediction performance. It validates the rationality of DGS_LWA-LLS.

Finally, DGS_LWA-LLS is compared with other granularity selection methods. For DGS_LWA-LLS, granularity selection is considered with respect to specific sample. In order to validate the

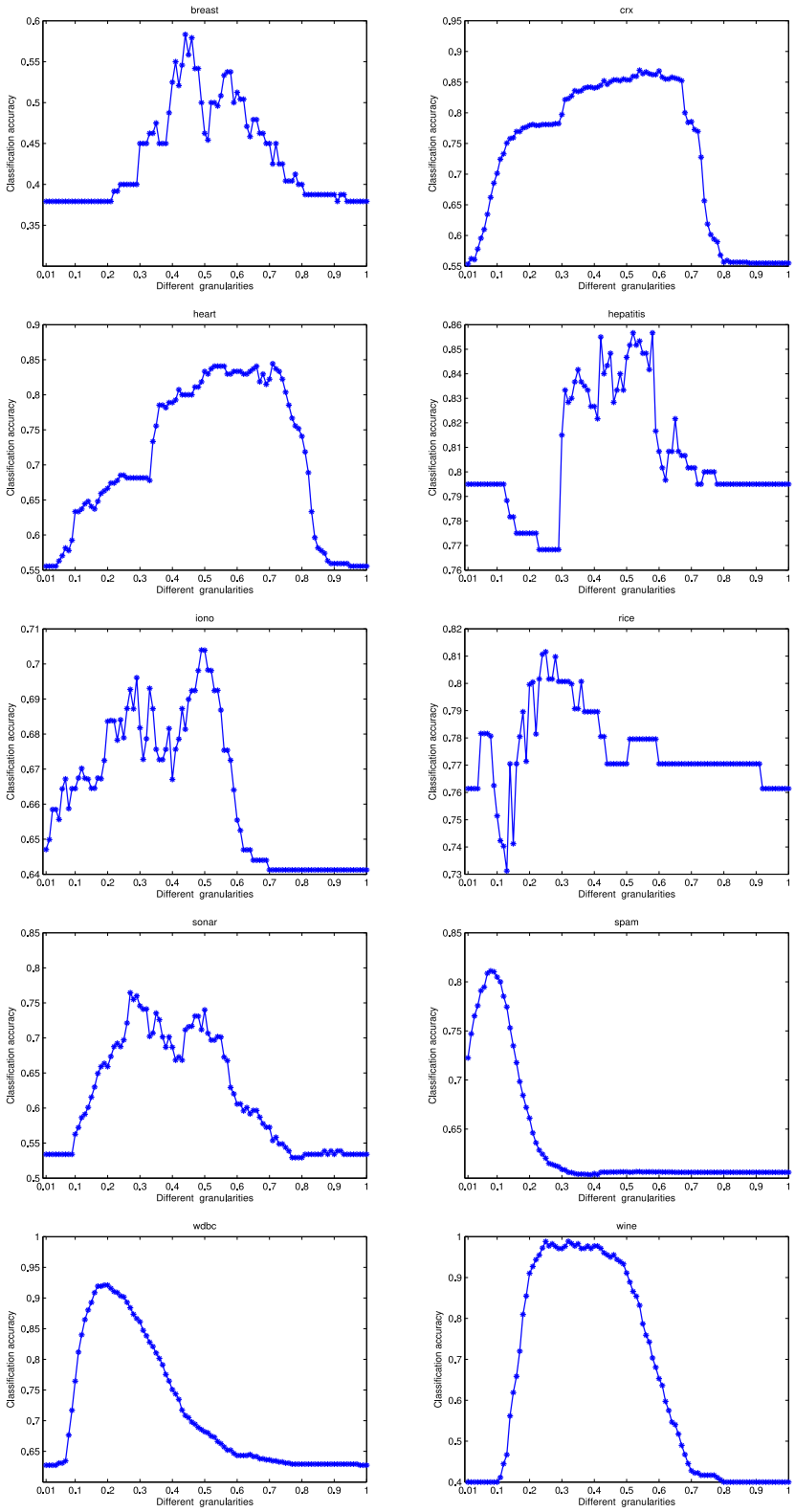

Fig. 6. Classification accuracies with different granularities.

effectiveness and superiority of dynamic granularity selection, the other granularity selection methods used for comparison includes GSC_MD, best fixed granularity (BFG) and Oracle.

(1) GSC_MD. As introduced in Section 2, GSC_MD [24] selects granularity with respect to specific data set, and different granularities will be selected for different data sets.

(2) BFG. As shown in Fig. 6, there is no granularity which can induce the best classification performance in every data set. Thus we select a granularity which can induce the best average classification accuracy. Note that the fixed granularity is selected based on test set. In other words, the selected granularity is the best fixed granularity. In fact, it is the theoretical upper bound

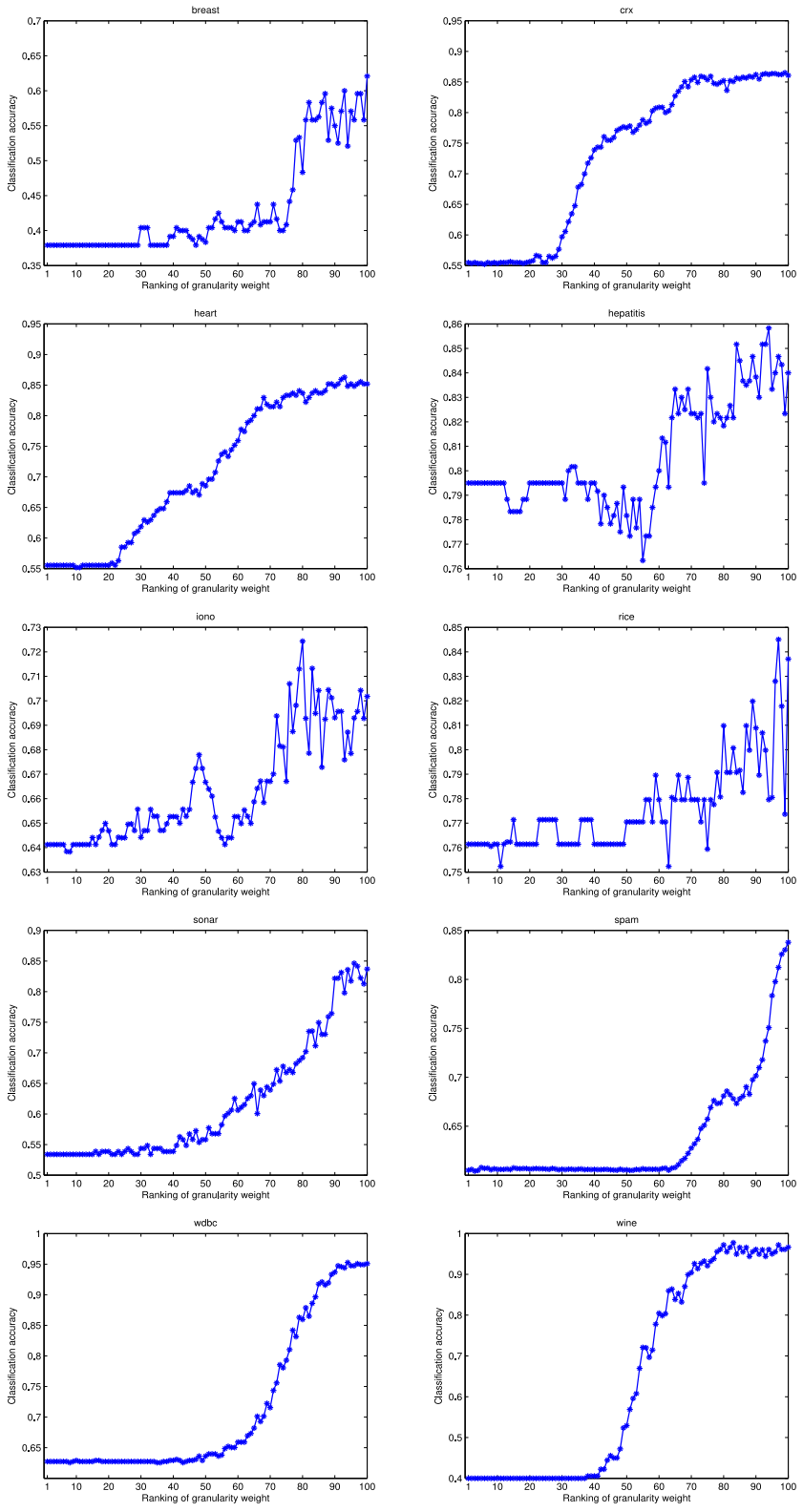

Fig. 7. Variation of classification accuracies with granularity weight.

when selecting a fixed granularity based on empirical values [35,36,39,40].

(3) Oracle. It is refer to the ideal granularity selector that always chooses the granularity, if any, with the correct classification. Oracle selects different granularities with respect to different samples, thus it is a kind of dynamic granularity selection method. In fact, Oracle is the theoretical upper bound for all granularity selection method with respect to classification performance.

For each granularity selection method, it is conducted based on a collection of granularities in which the granularity varies

Table 2
Performance comparison with other granularity selection methods.

Data Set	DGS_LWA-LLS	GSC_MD	BFG	Oracle
breast	62.08 ± 8.44	51.67 ± 9.86	58.83 ± 11.79	95.00 ± 8.74
crx	86.08 ± 15.51	84.49 ± 14.93	84.63 ± 15.36	96.55 ± 3.82
heart	85.19 ± 6.05	81.48 ± 6.30	80.00 ± 7.65	94.44 ± 5.01
hepatitis	84.00 ± 9.00	82.83 ± 9.88	84.33 ± 8.32	94.67 ± 5.26
iono	70.18 ± 4.93	68.38 ± 3.67	68.14 ± 3.17	91.28 ± 4.89
rice	83.71 ± 11.08	77.43 ± 11.73	77.05 ± 3.84	95.27 ± 6.58
sonar	83.69 ± 7.85	74.45 ± 11.91	71.14 ± 12.50	97.12 ± 4.61
spam	83.79 ± 3.66	78.98 ± 3.49	60.60 ± 0.12	96.46 ± 1.32
wdbc	95.09 ± 2.95	91.23 ± 4.29	70.85 ± 4.33	98.60 ± 1.61
wine	96.67 ± 3.88	98.89 ± 2.34	95.56 ± 4.38	100 ± 0.00
Average	83.05	78.98	75.06	95.94

from 0.01 to 1 with the step 0.01. In the experiments, the data set is randomly split into 10 subsets. In each run, 9 subsets are used for training set, 1 for testing set. The average generalization accuracy and the standard deviation over 10 runs are calculated to evaluate the classification performance of the method. Table 2 summarizes the classification performances of DGS_LWA-LLS, GSC_MD, BFG and Oracle. Some conclusions can be obtained:

(1) For best fixed granularity, the classification performances vary greatly with different classification tasks. Although it can get competitive classification results on some data sets (such as hepatitis), but its overall performance (the average generalization accuracy) is worse than the other methods. Thus it is not a good solution to use a fixed granularity.

(2) In general, GSC_MD is superior to BFG, and DGS_LWA-LLS performs better than GSC_MD and BFG. It indicates that, for a given classification task, selecting a proper granularity is necessary and selecting different granularities for different samples can further improve the classification performance.

(3) Comparing DGS_LWA-LLS with Oracle (the ideal dynamic granularity selection method), we can see that dynamic granularity selection still has great potential and needs to be further explored.

4.2. Multi-granularity subspaces based classification

Different granularity subspaces can be induced with different granularities. Here neighborhood dependency based attribute reduction (Definition 5) is employed to obtain different granularity subspaces with different neighborhood granularities. In the experiments, a fast forward heterogeneous attribute reduction algorithm called F2HARNRS [35] is utilized, and the neighborhood granularity δ varies from 0.01 to 1 with the step 0.01. In addition to 10 UCI data sets (Table 1), face recognition is considered in this subsection. Two face databases (ORL and Yale) are added to further evaluate the proposed method in multi-granularity subspaces. Fig. 8 shows the faces of the first person in the two face database. Each face image contains 32×32 pixels and is represented by a 1024-dimensional vector in image space. Detailed descriptions of these face databases are as follows:

(1) The ORL Face Database [53] contains 400 grayscale images of 40 individuals. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open or closed eyes, smiling or not smiling) and facial details (glasses or no glasses).

(2) The Yale Face Database [54] contains 165 grayscale images of 15 individuals. The images demonstrate variations in lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised, and wink), and with or without glasses.

In the experiments, we first discuss the influence of granularity on the classification performance and the dimension of granularity subspace. And then the rationality and effectiveness of DGS_LWA-LLS is validated in multi-granularity subspaces.

Fig. 9 shows the classification accuracies (the blue curves) and the numbers of attributes (the green curves) in different granularity subspaces. It can be observed that the classification performances and the dimension of granularity subspace first increase and then decrease. We can see that it is necessary to choose a proper granularity to obtain good classification performance.

Fig. 10 shows the classification accuracies with different granularity subspaces weights. Similar to Fig. 7, we can see the trend that larger weight usually means better prediction performance. It validates the rationality of DGS_LWA-LLS in multi-granularity subspaces. Table 3 summarizes the classification performances of DGS_LWA-LLS, GSC_MD, BFG and Oracle in multi-granularity subspaces. Similar to Table 2, we can find that: selecting a proper granularity (GSC_MD) is better than using fixed granularity (BFG) and selecting different granularities for different samples (DGS_LWA-LLS) can further improve the classification performance. Besides, we can also see, from Oracle, that dynamic granularity selection has great potential and needs to be further explored in multi-granularity subspaces.

5. Conclusions and future work

Granularity selection plays a crucial role in the utilization of granules when solving problem. It is a key problem in granular computing. In this work, we deliver a preliminary study to dynamic granularity selection and obtain some interesting and promising results. A granularity selection algorithm called DGS_LWA-LLS is given. Two indices, denoted as local weighted accuracy and local likelihood ratio, are introduced in DGS_LWA-LLS. They evaluate the quality of granularity with respect to specific sample. Local weighted accuracy can be considered as a kind of empirical risk in a granule, and local likelihood ratio reflects the locality of sample distribution in a granule. Experimental results validate the rationality and effectiveness of DGS_LWA-LLS. It can be concluded that selecting a proper granularity is better than using fixed granularity and selecting different granularities for different samples can further improve the classification performance.

Future studies and developments include, but are not limited to:

(1) In the future, we will design other indices to evaluate the quality of granularity with respect to specific sample and use them in dynamic granularity selection.

(2) Different granularity can provide complementary information. In the future, we will study the internal relationship between different granularities and explore dynamic granular fusion method.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106087.

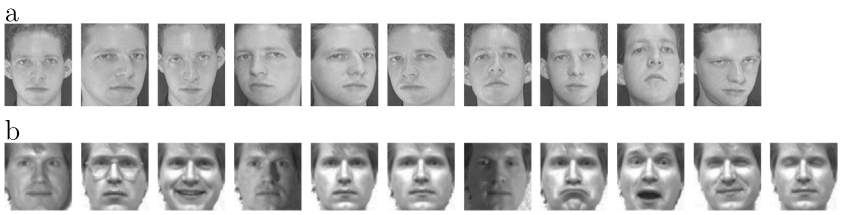

Fig. 8. (a) All faces of the first person in the ORL database; (b) All faces of the first person in the Yale database.

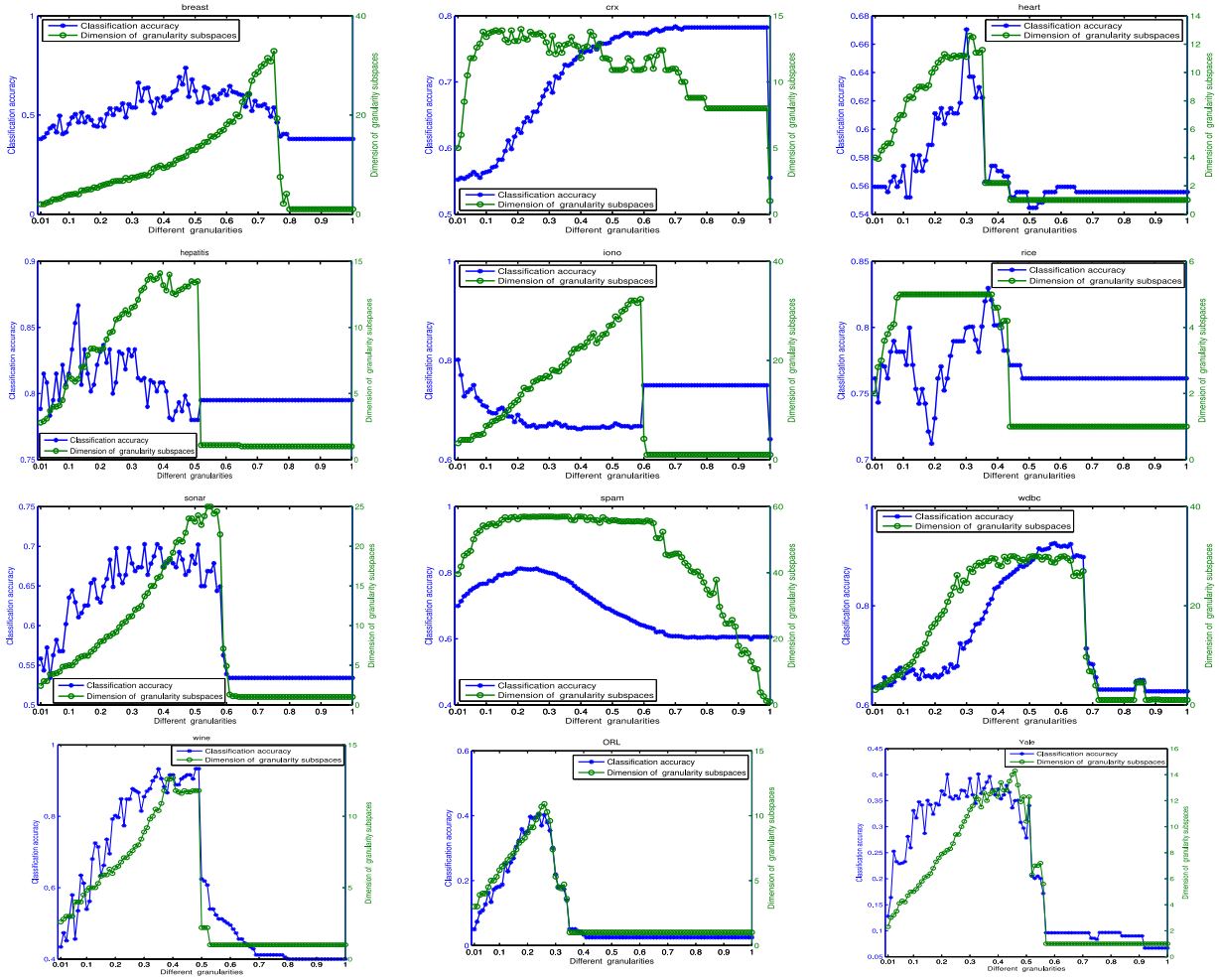

Fig. 9. Classification accuracies and numbers of attributes in different granularity subspaces.

Table 3
Performance comparison in multi-granularity subspaces.

Data set	DGS_LWA-LLS	GSC_MD	BFG	Oracle
breast	78.75 ± 8.44	67.08 ± 19.29	73.75 ± 13.76	100.00 ± 0.00
crx	76.79 ± 7.05	77.52 ± 6.61	75.79 ± 5.26	83.76 ± 2.56
heart	66.30 ± 11.64	66.67 ± 5.24	55.56 ± 3.49	82.96 ± 8.94
hepatitis	86.33 ± 8.95	84.67 ± 6.32	79.83 ± 4.19	96.17 ± 4.58
iono	83.62 ± 9.27	80.16 ± 6.70	66.72 ± 4.32	90.72 ± 5.83
rice	79.96 ± 15.50	77.16 ± 5.90	77.14 ± 5.65	93.45 ± 7.57
sonar	76.93 ± 4.34	69.76 ± 5.70	66.40 ± 6.45	88.93 ± 5.56
spam	82.57 ± 4.12	79.53 ± 3.19	69.88 ± 2.20	97.15 ± 1.08
wdbc	94.39 ± 3.77	92.46 ± 4.05	88.04 ± 2.88	97.19 ± 2.64
wine	96.11 ± 4.57	93.33 ± 9.00	90.49 ± 13.60	99.44 ± 1.76
ORL	74.42 ± 5.30	47.58 ± 7.01	49.92 ± 5.89	88.92 ± 1.42
Yale	68.72 ± 3.56	46.56 ± 7.95	52.67 ± 9.18	80.00 ± 5.67
Average	80.41	73.54	70.52	91.56

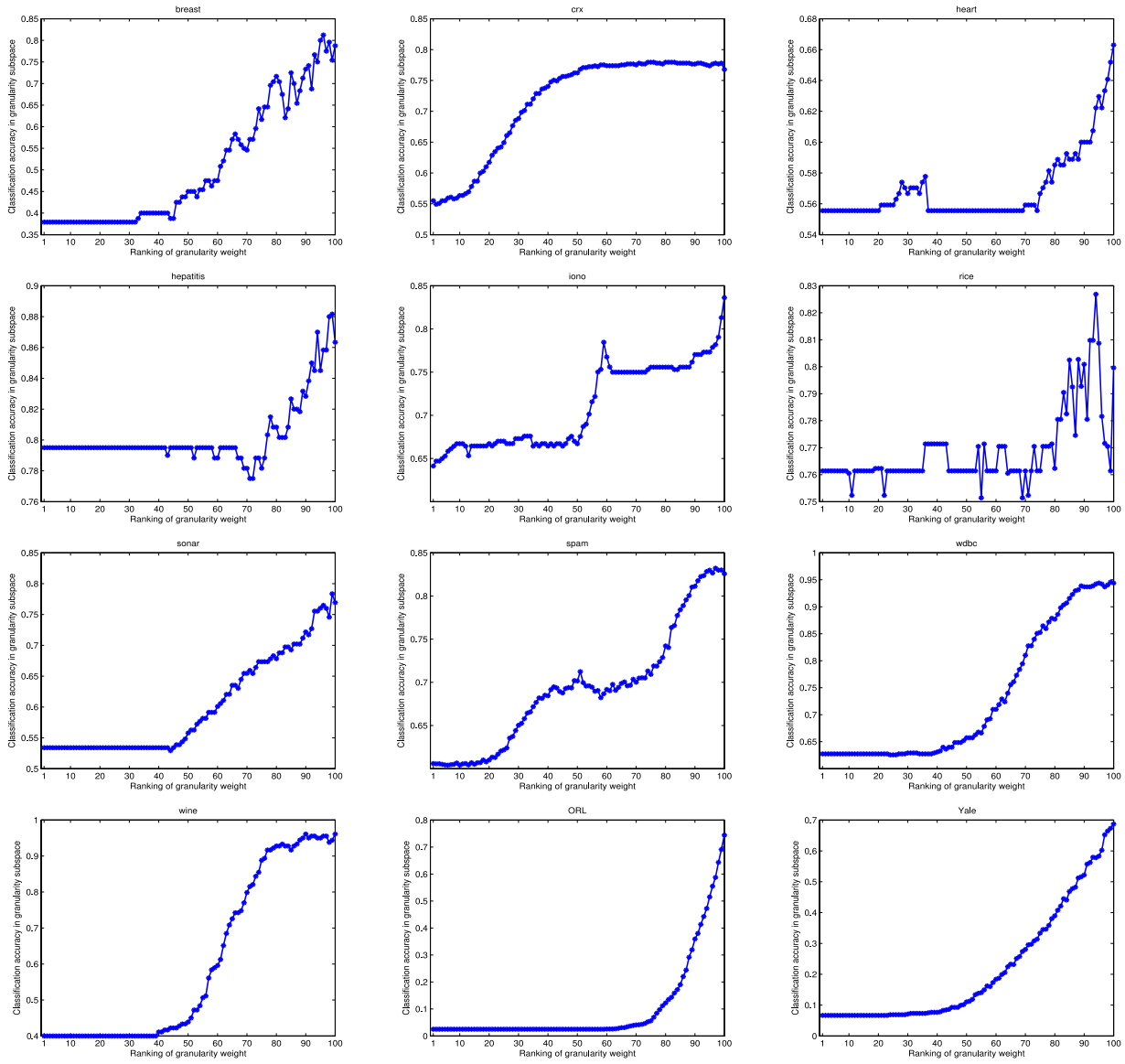

Fig. 10. Classification accuracies with different granularity subspaces weights.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 61502144, 61573127), the Natural Science Foundation of Hebei Province, China (Nos. F2018205196, F2019205295), the Natural Science Foundation of Higher Education Institutions of Hebei Province, China (No. BJ2019014), the Postdoctoral Advanced Programs of Hebei Province, China (No. B2016003013), the Doctor Science Foundation of Hebei Normal University (No. L2017B19).

References

[1] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (2) (1997) 111–127.
[2] W. Pedrycz, Granular Computing: An Emerging Paradigm, Physica-Verlag, Heidelberg.
[3] L.A. Zadeh, Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems, Soft Comput. 2 (1) (1998) 23–25.
[4] L.A. Zadeh, A new direction in fuzzy logic-toward a computational theory of perceptions, AI Mag. 22 (1) (2013) 73–84.
[5] G. Chiaselotti, T. Gentile, F. Infusino, Granular computing on information tables: Families of subsets and operators, Inform. Sci. 442–443 (2018) 72–102.
[6] V.J. Kok, C.S. Chan, Grcs: Granular computing-based crowd segmentation, IEEE Trans. Cybern. 47 (5) (2017) 1157–1168.
[7] Y.H. Qian, H.H. Cheng, J.T. Wang, J.Y. Liang, W. Pedrycz, C.Y. Dang, Grouping granular structures in human granulation intelligence, Inform. Sci. 382–383 (2017) 150–169.
[8] G. Wang, X. Ji, Granular computing with multiple granular layers for brain big data processing, Brain Inform. 1 (1–4) (2014) 1–10.
[9] C. Zhong, W. Pedrycz, D. Wang, L. Li, Z. Li, Granular data imputation: A framework of granular computing, Appl. Soft Comput. 46 (2016) 307–316.
[10] W. Pedrycz, Allocation of information granularity in optimization and decision-making models: Towards building the foundations of granular computing, European J. Oper. Res. 232 (1) (2014) 137–145.
[11] W. Pedrycz, W. Homenda, Building the fundamentals of granular computing: A principle of justifiable granularity, Appl. Soft Comput. 13 (10) (2013) 4209–4218.
[12] Y.Y. Yao, Granular computing: Basic issues and possible solutions, in: Proceedings of the 5th Joint Conference on Information Sciences, 2002, pp. 186–189.
[13] J.R. Hobbs, Granularity, in: Proceedings of the 9th International Joint Conference on Artificial Intelligence, 1985, pp. 432–435.
[14] M. Niepert, M. Gyssens, B. Sayrafi, D.V. Gucht, On the conditional independence implication problem: A lattice-theoretic approach, Artificial Intelligence 202 (9) (2013) 29–51.

[15] M.J. Wierman, Measuring uncertainty in rough set theory, Int. J. Gen. Syst. 28 (4–5) (1999) 283–297.
[16] J.M. Ma, W.X. Zhang, Y. Leung, X.X. Song, Granular computing and dual galois connection, Inform. Sci. 177 (23) (2007) 5365–5377.
[17] G.P. Lin, J.Y. Liang, Y.H. Qian, Multigranulation rough sets: From partition to covering, Inform. Sci. 241 (12) (2013) 101–118.
[18] W. Pedrycz, Granular Computing: Analysis and Design of Intelligent Systems, CRC Press, 2013.
[19] W.Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inform. Sci. 181 (18) (2011) 3878–3897.
[20] C. Hao, J.H. Li, M. Fan, W.Q. Liu, E.C.C. Tsang, Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions, Inform. Sci. 415 (2017) 213–232.
[21] F. Li, B.Q. Hu, A new approach of optimal scale selection to multi-scale decision tables, Inform. Sci. 381 (2017) 193–208.
[22] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Internat. J. Approx. Reason. 54 (8) (2013) 1107–1129.
[23] J.P. Xie, M.H. Yang, J.H. Li, Z. Zheng, Rule acquisition and optimal scale selection in multi-scale formal decision contexts and their applications to smart city, Future Gener. Comput. Syst. 73 (1) (2017) 1–30.
[24] P.F. Zhu, Q.H. Hu, Adaptive neighborhood granularity selection and combination based on margin distribution optimization, Inform. Sci. 249 (2013) 1–12.
[25] H. Zhao, W. Zhu, Optimal cost-sensitive granularization based on rough sets for variable costs, Knowl.-Based Syst. 65 (4) (2014) 72–82.
[26] H. Zhao, P. Wang, Q.H. Hu, Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence, Inform. Sci. 366 (2016) 134–149.
[27] Y. Liu, S.Z. Liao, Granularity selection for cross-validation of svm, Inform. Sci. 378 (2017) 475–483.
[28] W.H. Xu, W.T. Li, X.T. Zhang, Generalized multigranulation rough sets and optimal granularity selection, Granul. Comput. 2 (4) (2017) 1–18.
[29] Y.Y. Yao, Granular computing using neighborhood systems, in: Advances in Soft Computing, Springer, London, 1999, pp. 539–553.
[30] Z. Pawlak, A. Skowron, Rudiments of rough sets, Inform. Sci. 177 (1) (2007) 3–27.
[31] M. Kryszkiewicz, Rough set approach to incomplete information systems, Inform. Sci. 112 (1–4) (1998) 39–49.
[32] Y.M. Chen, Z.J. Zhang, J.Z. Zheng, Y. Ma, Y. Xue, Gene selection for tumor classification using neighborhood rough sets and entropy measures, J. Biomed. Inform. 67 (2017) 59–68.
[33] S.U. Kumar, H.H. Inbarani, Pso-based feature selection and neighborhood rough set-based classification for bci multiclass motor imagery task, Neural Comput. Appl. 28 (11) (2017) 3239–3258.
[34] B.Y. Li, J.M. Xiao, X.H. Wang, Feature reduction for power system transient stability assessment based on neighborhood rough set and discernibility matrix, Energies 11 (1) (2018) 1–19.
[35] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inform. Sci. 178 (18) (2008) 3577–3594.
[36] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2) (2008) 866–876.
[37] D.R. Wilson, T.R. Martinez, Improved heterogeneous distance functions, J. Artificial Intelligence Res. 6 (1) (1997) 1–34.
[38] T. Cover, Nearest neighbour pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27.
[39] Q.H. Hu, L. Zhang, D. Zhang, W. Pan, S. An, W. Pedrycz, Measuring relevance between discrete and continuous features based on neighborhood mutual information, Expert Syst. Appl. 38 (9) (2011) 10737–10750.
[40] C.Z. Wang, Q.H. Hu, X.Z. Wang, D.G. Chen, Y.H. Qian, D. Zhe, Feature selection based on neighborhood discrimination index, IEEE Trans. Neural Netw. Learn. Syst. 29 (7) (2017) 2986–2999.
[41] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough sets, Rough sets, Int. J. Parallel Program. 11 (5) (1982) 341–356.
[42] J. Dai, Q. Hu, J. Zhang, H. Hu, N. Zheng, Attribute selection for partially labeled categorical data by rough set approach, IEEE Trans. Cybern. 47 (9) (2017) 2460–2471.
[43] A.K. Das, S. Sengupta, S. Bhattacharyya, A group incremental feature selection for classification using rough set theory based genetic algorithm, Appl. Soft Comput. 65 (2018) 400–411.
[44] Y. Yang, D. Chen, H. Wang, Active sample selection based incremental algorithm for attribute reduction with rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2017) 825–838.
[45] T.Y. Lin, Granulation and nearest neighborhoods: Rough set approach, in: Granular Computing, 2001, pp. 125–142.
[46] T.Y. Lin, Neighborhood systems: mathematical models of information granulations, in: IEEE International Conference on Systems, Man and Cybernetics, 2003, pp. 3188–3193.
[47] C.Z. Wang, Y.L. Qi, M.W. Shao, Q.H. Hu, D.G. Chen, Y.H. Qian, Y.J. Lin, A fitting model for feature selection with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2017) 741–753.
[48] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (4) (2008) 294–304.
[49] W.Z. Zhu, G.Q. Si, Y.B. Zhang, J.C. Wang, Neighborhood effective information ratio for hybrid feature subset evaluation and selection, Neurocomputing 99 (2013) 25–37.
[50] Q.H. Hu, X.J. Che, L. Zhang, D.R. Yu, Feature evaluation and selection based on neighborhood soft margin, Neurocomputing 73 (10–12) (2010) 2114–2124.
[51] J.G. Kalbfleisch, Probability and statistical inference. vol. 2, in: Statistical inference, in: Springer Texts in Statistics, 141, (2) 1985, pp. 391–392.
[52] C.L. Blake, C.J. Merz, UCI repository of machine learning databases, https://archive.ics.uci.edu/ml/index.php.
[53] ORL Face Database, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.
[54] Yale Face Database, http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html.