

Article

Neighborhood Granule Classifiers

Hongbo Jiang¹ and Yumin Chen^{2,*}

¹ School of Economics and Management, Xiamen University of Technology, Xiamen 361024, China; hbjiang@xmut.edu.cn

² School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

* Correspondence: cym0620@163.com

Received: 11 October 2018; Accepted: 11 December 2018; Published: 17 December 2018



Abstract: Classifiers are divided into linear and nonlinear classifiers. The linear classifiers are built on a basis of some hyper planes. The nonlinear classifiers are mainly neural networks. In this paper, we propose a novel neighborhood granule classifier based on a concept of granular structure and neighborhood granules of datasets. By introducing a neighborhood rough set model, the condition features and decision features of classification systems are respectively granulated to form some condition neighborhood granules and decision neighborhood granules. These neighborhood granules are sets; thus, their calculations are intersection and union operations of sets. A condition neighborhood granule and a decision neighborhood granule form a granular rule, and the collection of granular rules constitutes a granular rule library. Furthermore, we propose two kinds of distance and similarity metrics to measure granules, which are used for the searching and matching of granules. Thus, we design a granule classifier by the similarity metric. Finally, we use the granule classifier proposed in this paper for a classification test with UCI datasets. The theoretical analysis and experiments show that the proposed granule classifier achieves a better classification performance under an appropriate neighborhood granulation parameter.

Keywords: neighborhood granule classifiers; neighborhood rough sets; granular structures; granular distances; granular computing

1. Introduction

Classification is a supervised learning method. It establishes a mapping from a feature space onto classes so that things can be classified and identified. Many classification problems are encountered in various fields, including graphic images [1], medicine [2], economics [3], finance [4], cyber security [5], etc. In order to improve the performance and accuracy of classification, many classification methods have been proposed. Classification methods mainly focus on statistical analysis [6,7], neural networks [8–10], rule reasoning [11,12], etc.

In the classification methods of statistical analysis theory, scholars have developed many classification techniques and models, such as decision trees [13], random forests [14], Bayesian models [15], logistic regression models [16], k-Nearest Neighbor classification [17,18], etc. These classical statistical classification methods are built on the basis of the Bayesian decision theory [19]. The theoretical premise is that the prior probability should be known. However, the prior probability is often difficult to be truly obtained.

Many scholars turned to study classification technologies involving neural networks. Vapnik et al. proposed the support vector machine (SVM) [20–22], which is a shallow neural network model. Based on the theory of structural risk minimization, the optimal hyper plane is constructed in the feature space, which makes the classifier globally optimized and achieves excellent classification performance in a small sample environment. SVM has been widely applied in face recognition [23,24],

image segmentation [25], pedestrian detection [26], gene selection [27,28], and so on. In recent years, the shallow neural networks have developed into deep neural networks. Hinton et al. proposed the concept of deep learning in 2006 [29,30], which solved the optimization problems caused by deep structures to a certain extent. Deep learning has achieved a great success in computer vision [31–33], speech recognition [34], and natural language processing [35]. However, with the number of deep neural network layers and the number of neurons in each layer increasing, the parameters of the inter-layer connections become very large, which result in being slow in the global convergence speed. The neural network is a black box and lacks theoretical explanations and guidance. Parameter tuning often relies on experience.

It is well known that fuzzy logic can deal with uncertainty and fuzziness [36,37]. The classification rules derived from fuzzy logic have good interpretability. The use of rules like “if-then” provides a good insight into the structure of the classifier, which makes the classification results have good interpretability. However, the formation of the rule base is inseparable from the domain knowledge of experts, and the speed of searching and matching in the rule base is slow, while the learning ability is weak.

The classifiers are divided into linear classifiers and nonlinear classifiers. The linear classifiers are established on the basis of the hyper planes. The nonlinear classifiers are built on neural networks. In this paper, we propose a new classifier model based on the set theory and the neighborhood information granulation. The concept of information granulation was introduced by Zadeh [38], and the concept of granular computing was proposed by T.Y. Lin [39,40]. Yao studied the neighborhood systems and neighborhood granular computing [41–43], the applications of which were developed by Hu in neighborhood reduction and classification [44–46]. Morente-Molinera proposed a multigranular linguistic modeling with fuzzy entropy for supervised classification learning purposes [47]. Pedrycz developed a certain category of granule classifiers referred to as hyper box-driven classifiers [48–50]. We focus on the neighborhood granulation of classification systems and propose a granular structure, granular distance, and granular rule by a neighborhood granulation, so as to build a new granule classifier model. Since the granules are formed with sets, the granule classifier is a set-based classifier. Furthermore, we present the granular similarity for searching and matching in a granular rule base, design a granule classifier, and verify it experimentally. Theoretical analysis and experimental results show that the proposed classifier can achieve a better classification performance under a suitable granulating parameter.

The paper is structured as follows. First, in Section 2, we introduce the neighborhood granulation of classification systems. Then, we propose a new neighborhood granule classifier in Section 3. In Section 4, we present experimental results. Conclusions and future works are covered in Section 5.

2. Neighborhood Granulation of a Classification System

The rough set theory (RST) [51] proposed by Pawlak, a mathematician of Poland, is a widely used model in classification systems. In a classification system, the data with real values need to be discretized, but the process of discretization easily leads to a loss of classification information. Aiming at the limitation of Pawlak RST, a neighborhood RST is introduced to granulate these real-value data. In RST, an equivalence class is regarded as a granule. As for neighborhood RST, neighborhood granules are constructed.

Suppose $CS = (U, A, D, \delta)$ is a classification system, where $U = \{x_1, x_2, \dots, x_n\}$ is a set of samples or objects; $A = \{a_1, a_2, \dots, a_m\}$ is a set of condition features or attributes; $D = \{d\}$ represents a decision feature or attribute; $\delta \in [0, 1]$ is a neighborhood parameter [44]. The data of samples in the condition feature set A are real values, and the data of samples in the decision feature D are symbolic or discrete values.

Suppose $CS = (U, A, D, \delta)$ is a classification system. For any sample $x, y \in U$ and a feature subset $P \subseteq A$, where $P = \{a_1, a_2, \dots, a_m\}$, a distance function on the feature subset P is defined as [44]:

$$\Delta_P(x, y) = \left(\sum_{i=1}^m (v(x, a_i) - v(y, a_i))^s \right)^{1/s}, \tag{1}$$

where $v(x, a_i)$ represents the value of the sample x on a feature a_i . If $s = 1$, it is named the Manhattan distance, and if $s = 2$, it is named the Euclidean distance.

Suppose $CS = (U, A, D, \delta)$ is a classification system. For any sample $x \in U$ and a condition feature subset $P \subseteq A$, the δ neighborhood granule of x on P is formed by:

$$g_P^\delta(x) = \{y | x, y \in U, \Delta_P(x, y) \leq \delta\}, \tag{2}$$

which is called a process of neighborhood granulation.

The data on decision features in classification systems are symbolic values. The samples are divided into equivalent classes on the decision features and the neighborhood parameter $\delta = 0$. Therefore, these samples are granulated into equivalent granules, which are called decision equivalent granules, and they are expressed as:

$$g_D^0(x) = \{y | x, y \in U, \Delta_D(x, y) = 0\}. \tag{3}$$

It is obvious that a neighborhood granule is degenerated into an equivalent granule when the neighborhood parameter δ is zero. Therefore, the equivalent granule is a special neighborhood granule. In a classification system, for any sample $x \in U$ and a condition feature subset $P \subseteq A$, it determines a neighborhood relation, which is defined as:

$$NR_\delta(P) = \{(x, y) \in U \times U | \Delta_P(x, y) \leq \delta\}. \tag{4}$$

The $U/NR_\delta(P)$ is called a covering of the domain U , which is induced by a neighborhood relation, and the covering is a collection of neighborhood granules. The neighborhood relation is a kind of similarity relation that satisfies reflexive and symmetric properties, rather than the equivalence relation. When a neighborhood parameter δ is equal to zero, the neighborhood relation is degenerated into an equivalence relation. Therefore, the equivalence relation is a special neighborhood relation. The rough sets based on neighborhood relations can handle the real-value data, while the rough sets based on equivalence relations can only deal with the symbolic data.

A classification system is $CS = (U, A, D, \delta)$. For any sample $x \in U$ and a feature subset $P \subseteq A$, let $g_P^\delta(x)$ be a neighborhood granule, then the size of granule $g_P^\delta(x)$ is defined as:

$$Size(g_P^\delta(x)) = \frac{|g_P^\delta(x)|}{|U|}, \tag{5}$$

where $|\cdot|$ represents the cardinality of a set. It is easy to know that the size of the neighborhood granule holds the property: $\frac{1}{|U|} \leq Size(g_P^\delta(x)) \leq 1$.

Example 1. A classification system $CS = (U, A, D, \delta)$ is shown in Table 1. Suppose $U = \{x_1, x_2, x_3, x_4\}$ is the sample set, $A = \{a, b, c\}$ is the feature set, and the neighborhood granulation parameter $\delta = 0.1$. This example uses the Euclidean distance for neighborhood granulation.

Table 1. A classification system.

U	a	b	c	$\Rightarrow d$
x_1	0.1	0.2	0.1	1
x_2	0.2	0.5	0.2	1
x_3	0.3	0.3	0.3	0
x_4	0.7	0.1	0.3	0

The domain is $U = \{x_1, x_2, x_3, x_4\}$. We can granulate the samples on feature subset $P = \{a\}$. These neighborhood granules are $g_1 = g_P^{0.1}(x_1) = \{x_1, x_2\}$, $g_2 = g_P^{0.1}(x_2) = \{x_1, x_2, x_3\}$, $g_3 = g_P^{0.1}(x_3) = \{x_2, x_3\}$, and $g_4 = g_P^{0.1}(x_4) = \{x_4\}$.

The sizes of condition granules are calculated as: $Size(g_1) = Size(g_3) = 1/2$, $Size(g_2) = 3/4$, $Size(g_4) = 1/4$.

According to the neighborhood granulation of feature subset $Q = \{a, b\}$, the condition neighborhood granules are: $g_5 = g_Q^{0.1}(x_1) = \{x_1\}$, $g_6 = g_Q^{0.1}(x_2) = \{x_2\}$, $g_7 = g_Q^{0.1}(x_3) = \{x_3\}$, $g_8 = g_Q^{0.1}(x_4) = \{x_4\}$.

The sizes of condition granules are calculated as: $Size(g_5) = Size(g_6) = Size(g_7) = Size(g_8) = 1/4$. It is easy to know that if $P \subseteq Q$, then $Size(g_P^{0.1}(x_1)) > Size(g_Q^{0.1}(x_1))$. The decision equivalent granules are $d_1 = g_d^0(x_1) = \{x_1, x_2\}$, $d_2 = g_d^0(x_2) = \{x_1, x_2\}$, $d_3 = g_d^0(x_3) = \{x_3, x_4\}$, $d_4 = g_d^0(x_4) = \{x_3, x_4\}$.

3. The Model of Neighborhood Granule Classifiers

A granular rule is defined in this paper, which is constructed by a condition granule and a decision granule, being formed by a neighborhood granulation of a sample. The condition granule is a precursor of the granular rule, while the decision granule is a consequence of the granular rule. The granular rules of all the samples constitute a granular rule base. Then, two kinds of granular distances and granular similarity measures are defined so that the searching and matching of granules can be performed. A test sample is granulated into a condition granule as a test granule. The test granule is matched with the precursor of a granular rule in a granular rule base. The label of the most numerous classes in the matched granular rules is the predictive label of the test sample.

3.1. Neighborhood Granular Structure

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A, \forall x \in U$, there exists a neighborhood granule $g_B^\delta(x)$ on B . The set of all neighborhood granules on B is called the neighborhood granular swarm, which is defined as:

$$G_B^\delta = \{g_B^\delta(x) | \forall x \in U\}. \tag{6}$$

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A$, let G_B^δ be a neighborhood granular swarm; the size of neighborhood granular swarm G_B^δ is defined as:

$$GM(G_B^\delta) = \frac{1}{|U|} \sum_{i=1}^{|U|} Size(g_B^\delta(x_i)) = \frac{1}{|U|^2} \sum_{i=1}^{|U|} |g_B^\delta(x_i)|. \tag{7}$$

It is easy to know that the size of neighborhood granular swarm holds the property: $\frac{1}{|U|} \leq GM(G_B^\delta) \leq 1$.

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A, \forall x \in U$, there exists a neighborhood granule $g_B^\delta(x)$ on B and a decision equivalent granule $g_D^0(x)$ on D . A neighborhood granular rule is defined as an ordered pair: $r_B(x) = \langle g_B^\delta(x), g_D^0(x) \rangle$. The neighborhood granular rule base is defined as: $K_B = \{r_B(x) | \forall x \in U\}$.

The samples of a classification system are granulated into neighborhood granules by neighborhood parameters, and the set of neighborhood granules constitutes a neighborhood granular swarm. Condition granules and decision granules form a granular rule, and the collection of granular rules constitutes a granular rule base. Therefore, the process of classification can be transformed into the process of reasoning, searching, and matching in the granular rule base.

3.2. Granular Distance between Neighborhood Granules

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A$ and $\forall x, y \in U$, suppose $p = g_B^\delta(x)$, $q = g_B^\delta(y)$ are two neighborhood granules, then the relative distance between two neighborhood granules is defined as:

$$d(p, q) = \frac{|p \oplus q|}{|p \vee q|} = \frac{|(g_B^\delta(x) \cup g_B^\delta(y)) - (g_B^\delta(x) \cap g_B^\delta(y))|}{|g_B^\delta(x) \cup g_B^\delta(y)|} \tag{8}$$

Example 2. There are two neighborhood granules such as $g_1 = \{x_1, x_2\}$ and $g_2 = \{x_1, x_2, x_3\}$, respectively. We can obtain their relative granular distance: $d(g_1, g_2) = \frac{|g_1 \oplus g_2|}{|g_1 \vee g_2|} = \frac{|(\{x_1, x_2\} \cup \{x_1, x_2, x_3\}) - (\{x_1, x_2\} \cap \{x_1, x_2, x_3\})|}{|\{x_1, x_2\} \cup \{x_1, x_2, x_3\}|} = \frac{|\{x_3\}|}{|\{x_1, x_2, x_3\}|} = 1/3$.

Proposition 1. The relative distance between two neighborhood granules is a distance metric, which satisfies the following three properties:

- (1) $d(p, q) \geq 0$, non-negativity;
- (2) $d(p, q) = d(q, p)$, symmetry;
- (3) $d(p, k) + d(k, q) \geq d(p, q)$, triangle inequality.

Proof. The two properties of (1) and (2) are easily proven by the definition of the relative distance of neighborhood granules. The following proves the property (3).

Given any a, b, c , let $a \geq b > 0$, $c \geq 0$, then $\frac{b}{a} \leq \frac{b+c}{a+c}$. According to the definition of the relative distance between neighborhood granules, we can know $d(p, k) + d(k, q) - d(p, q) = \frac{|p \cup k| - |p \cap k|}{|p \cup k|} + \frac{|k \cup q| - |k \cap q|}{|k \cup q|} - \frac{|p \cup q| - |p \cap q|}{|p \cup q|} = 1 - \frac{|p \cap k|}{|p \cup k|} - \frac{|k \cap q|}{|k \cup q|} + \frac{|p \cap q|}{|p \cup q|} \geq 1 - \frac{|p \cap k| + (|q| - |q \cap p \cup k|)}{|p \cup k| + (|q| - |q \cap p \cup k|)} - \frac{|k \cap q| + (|p| - |p \cap k \cup q|)}{|k \cup q| + (|p| - |p \cap k \cup q|)} + \frac{|p \cap q|}{|p \cup k \cup q|} = 1 - \frac{|p \cap k| + |q| - (|p \cap q| + |k \cap q| - |p \cap k \cap q|)}{|p \cup k \cup q|} - \frac{|k \cap q| + |p| - (|p \cap k| + |p \cap q| - |p \cap k \cap q|)}{|k \cup q| + |p| - (|p \cap k| + |p \cap q| - |p \cap k \cap q|)} + \frac{|p \cap q|}{|p \cup k \cup q|} = 1 - \frac{|p| + |q| - |p \cap q|}{|p \cup k \cup q|} + \frac{2(|p \cap q| - |p \cap k \cap q|)}{|p \cup k \cup q|} \geq 0$. Therefore, $d(p, k) + d(k, q) - d(p, q) \geq 0$. That is $d(p, k) + d(k, q) \geq d(p, q)$. \square

Proposition 2. The granular relative distance of any two neighborhood granules satisfies: $0 \leq d(p, q) \leq 1$.

Proof. According to the definition of the relative distance of neighborhood granules, it is easy to prove.

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A$ and $\forall x, y \in U$, suppose $p = g_B^\delta(x)$, $q = g_B^\delta(y)$ are two neighborhood granules, then the absolute distance between two neighborhood granules is defined as:

$$h(p, q) = \frac{|p \oplus q|}{|U|} = \frac{|(g_B^\delta(x) \cup g_B^\delta(y)) - (g_B^\delta(x) \cap g_B^\delta(y))|}{|U|} \tag{9}$$

\square

Example 3. There are two neighborhood granules such as $g_1 = \{x_1, x_2\}$ and $g_2 = \{x_1, x_2, x_3\}$, respectively. The sample domain is $U = \{x_1, x_2, x_3, x_4\}$. We can obtain their absolute granular distance: $h(g_1, g_2) = \frac{|g_1 \oplus g_2|}{|U|} = \frac{|(\{x_1, x_2\} \cup \{x_1, x_2, x_3\}) - (\{x_1, x_2\} \cap \{x_1, x_2, x_3\})|}{|\{x_1, x_2, x_3, x_4\}|} = \frac{|\{x_3\}|}{|\{x_1, x_2, x_3, x_4\}|} = 1/4$.

Proposition 3. The absolute distance between two neighborhood granules satisfies: $0 \leq h(p, q) \leq 1$.

Proof. According to the definition of the absolute distance of neighborhood granules, it is easy to prove.

These granular distances of neighborhood granules can be used as the similarity measures of granules, which indicate the degree of similarity of neighborhood granules. The smaller the granular distance is, the greater the similarity is. On the contrary, the larger the granular distance is, the smaller the similarity is. \square

3.3. Neighborhood Granule Classifiers

Suppose $CS = (U, A, D, \delta)$ is a classification system. For $\forall B \subseteq A$ and $\forall x, y \in U$, suppose $p = g_B^\delta(x)$, $q = g_B^\delta(y)$ are two neighborhood granules, then the similarity measures between two neighborhood granules are defined as:

Relative similarity measure: $simd(p, q) = 1 - d(p, q)$.

Absolute similarity measure: $simh(p, q) = 1 - h(p, q)$.

Suppose $CS = (U, A, D, \delta)$ is a classification system. $S \subseteq U$ is a training sample set, and $T \subseteq U$ is a test sample set. For $\forall x \in S$, $r_A(x) = \langle g_A^\delta(x), g_D^0(x) \rangle$ is a granular rule, and $K_A = \{r_A(x) | \forall x \in S\}$ is a granular rule base. For $\forall y \in T$, $g_A^\delta(y)$ is a test neighborhood granule, which is granulated by the training sample set. Suppose $\eta \in [0, 1]$ is a similarity parameter, $\forall x \in S$, if $simd(g_A^\delta(y), g_A^\delta(x)) \geq \eta$ or $simh(g_A^\delta(y), g_A^\delta(x)) \geq \eta$, then a granular rule is matched from the granular rule base, and the decision granule of the granular rule is merged into a decision granule set. This matching process is a granular search and is expressed as follows:

Using the relative similarity measure: $Ld(y) = \{g_D^0(x) | \forall x \in S, simd(g_A^\delta(y), g_A^\delta(x)) \geq \eta\}$.

Using the absolute similarity measure: $Ld(y) = \{g_D^0(x) | \forall x \in S, simh(g_A^\delta(y), g_A^\delta(x)) \geq \eta\}$.

Suppose $CS = (U, A, D, \delta)$ is a classification system. $S \subseteq U$ is a training sample set, and $T \subseteq U$ is a test sample set. For $\forall y \in T$, $Ld(y)$ is the decision granule set of the test sample y , then the label of the test sample y is the most numerous granule label in $Ld(y)$.

4. Design of a Neighborhood Granule Classifier

A neighborhood granule is a set. The neighborhood granule classifier is a classifier with set operations. The classifier is divided into the granulation process and the classification process. In this section, we discuss the principle of the neighborhood granule classifier and give the algorithm of the neighborhood granule classification.

4.1. Granulation and Classification of Neighborhood Granules

The neighborhood granule classifier has only the granulation and the classification steps, and it has no training process. The granulation process includes: data preprocessing, partitioning of a training set and a test set, the training set granulated into a granular rule base, and the test set granulated into a set of test granules. The classification process includes: searching and matching in the granular rule base for test granules and determining the labels of test granules. The detailed granulation and classification processes are presented as follows:

1. Data preprocessing: deleting the data with missing values and normalizing the data with the interval $[0, 1]$.
2. Divided with a training set and a test set: 80% is the training set, and 20% is the test set.
3. The neighborhood granulation of the training set: a granular rule base is constructed according to the Euclidean distance and neighborhood parameters.
4. The neighborhood granulation of the test set: for a test sample, computing the Euclidean distance between the test sample and each training sample to form a test granule, and all the test samples are granulated into the test granular set.
5. Searching and matching for the test granule: setting the value of a similarity parameter, for the test granule, computing the similarity between the test granule and the condition granule of a rule in the granular rule base, getting all the granular rules whose similarities are larger than the

similarity threshold, and collecting the decision granules of matched granular rules to form a set of decision granules. Note: the elements of a decision granular set may be identical.

6. Judging the label of the test granule: The label of the test granule is the most numerous granule label in the decision granule set.

From the above analysis, it can be known that the neighborhood granule classifier is similar to the KNN classifier, and it is a lazy learning algorithm.

4.2. The Algorithm for Neighborhood Granule Classification

From the analysis of a structure of neighborhood granules, we propose the distance and similarity measures for neighborhood granules. Therefore, we can design a neighborhood granule classifier. The detailed neighborhood granule classification algorithm is described as follows:

Algorithm 1: Neighborhood granule classification (NGC).

Input: A training set $S = (U, A, D)$, a test sample t , a neighborhood parameter $\delta \in [0, 1]$, and a similarity parameter $\eta \in [0, 1]$.

Output: The *label* of a test sample t .

- (1) Normalization: $U \in [0, 1], t \in [0, 1]$.
 - (2) For each training sample $x \in U$
 - (3) Compute a neighborhood granule $g_A^\delta(x)$ by the neighborhood parameter δ ;
 - (4) Achieve an equivalence granule $g_D^0(x)$;
 - (5) Form a granular rule $r_A(x) = \langle g_A^\delta(x), g_D^0(x) \rangle$, and insert it into a granular rule base L ;
 - (6) End for.
 - (7) For the test sample t , granulate it as $g_A^\delta(t)$ according to the training set.
 - (8) For each granular rule $\langle g_A^\delta(x), g_D^0(x) \rangle \in L$
 - (9) Compute the granular distance $d(g_A^\delta(x), g_A^\delta(t))$;
 - (10) Compute the granular similarity degree $simd(g_A^\delta(x), g_A^\delta(t)) = 1 - d(g_A^\delta(x), g_A^\delta(t))$;
 - (11) If $simd(g_A^\delta(x), g_A^\delta(t)) \geq \eta$, then achieve the granular rule $\langle g_A^\delta(x), g_D^0(x) \rangle$, and insert the decision granule $g_D^0(x)$ into a target base T ;
 - (12) End for.
 - (13) The *label* of the test sample t is the label of a granule from T , and its quantity is the maximum.
 - (14) Return *label*.
-

In the NGC algorithm, the neighborhood granulation process is mainly involved. In Step 3, the time complexity of neighborhood granulation is $O(m * n)$, while the number of features is m , and that of training samples is n . Therefore, the time complexity of neighborhood granulation for training samples is $O(m * n^2)$ in Steps 2–6. Step 7 is the neighborhood granulation of the test set, and its time complexity is $O(m * t^2)$, while t is the number of test samples and $t < n$. Steps 8–14 have linear complexity. Therefore, in the worst case, the time complexity of the NGC algorithm is $O(m * n^2)$.

5. Experimental Analyses

This paper used four datasets from the UCI machine learning repository <http://archive.ics.uci.edu/ml/index.php> as experimental data sources, which are shown in Table 2.

Due to the different range of values of the datasets, the datasets needed to be normalized and preprocessed. Since values of neighborhood parameter varied between zero and one, we use the

max-min normalization method to ensure that all data could be converted into values between [0, 1]. The formula of max-min normalization is the following:

$$f(x_i) = \frac{x_i - x_{min}}{x_{max} - x_{min}}. \tag{10}$$

Table 2. Datasets used in the experiments.

Datasets	Samples	Features	Classes
Glass	214	9	6
Pima	768	8	2
Segmentation	210	19	7
Wine	178	13	3

All the classification experiments were implemented by a five-fold cross-validation on the four datasets. We split the dataset into five equally-sized subsets. Of the five subsets, a single subset was retained as the testing data, and the remaining four subsets were used as training data. The cross-validation process was then repeated five times, with each of the five subsets used exactly once as the testing data. The classification accuracy was the mean of five tests. In order to test the classification accuracy of our proposed granule classifier, the values of neighborhood granulation parameters were varied from 0.55–0.95 with an interval of 0.05, which involved a total of nine times for testing. The neighborhood granule classifier needed to set thresholds of granular similarity measure, which started from 0.55–0.95 by intervals of 0.05 with a total of nine changes. The granular similarity measures in these experiments used the absolute similarity measure. The experimental results are shown in Figures 1–4.

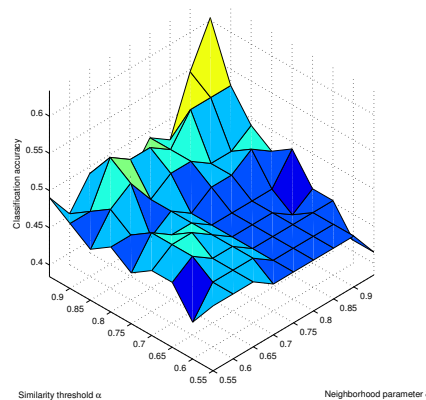


Figure 1. Classification accuracy for the glass dataset.

As can be seen from Figure 1, for the glass dataset, when the neighborhood parameter δ was 0.95 and the similarity threshold α was 0.95, the classification accuracy reached a maximum of 0.6038. The next values of classification accuracy in maximum were mainly distributed in the corners where granular similarity thresholds and neighborhood granulation parameters were both large. The smallest values of classification accuracy were mainly distributed on the plane with smaller granular similarity thresholds and larger neighborhood granular parameters.

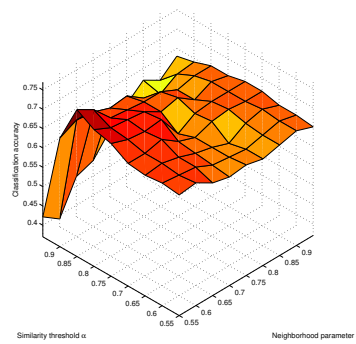


Figure 2. Classification accuracy for the pima dataset.

As can be seen from Figure 2, for the pima dataset, when the neighborhood parameter δ was 0.55 and the similarity threshold α was 0.85, the classification accuracy reached a maximum of 0.75. The second largest values of the classification accuracy were basically distributed on the whole plane. The minimum value of classification accuracy was 0.3906, and it was distributed at the point where (δ, α) was (0.60, 0.95).

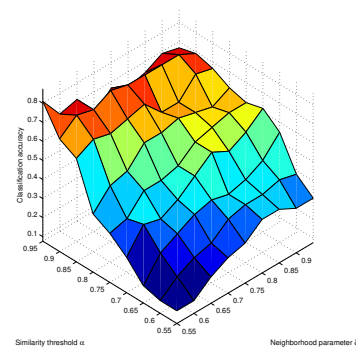


Figure 3. Classification accuracy for the segmentation dataset.

As can be seen from Figure 3, for the segmentation dataset, when the (δ, α) was (0.55, 0.95), the classification accuracy reached a maximum of 0.8095. The second largest values were mainly distributed on the band plane with larger similarity thresholds. The minimum values of the classification accuracy were mainly distributed on the plane where similarity thresholds and neighborhood parameters were small at the same time, and the sub-minimum values of the classification accuracy were mainly distributed on the band plane where the similarity threshold was small.

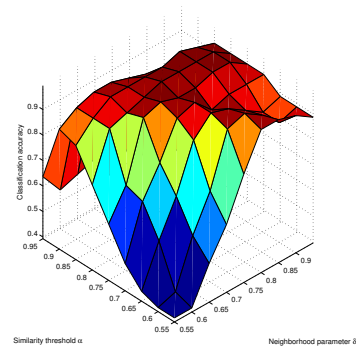


Figure 4. Classification accuracy for the wine dataset.

As can be seen from Figure 4, for the wine dataset, there were 11 maximum values of classification accuracy which were equal to 0.9773, mainly distributed on the plane between the neighborhood parameters ranging from 0.70–0.90 and the similarity thresholds ranging from 0.6–0.85. The minimum values of classification accuracy were mainly distributed in corners with neighborhood parameters being smaller and similarity thresholds being smaller at the same time.

From the above experiments, we can see that, when the dataset had a large number of categories, such as glass and segmentation, the classification accuracy was mainly related to the granular similarity threshold. The similarity threshold was larger, and the classification effect was better. When the dataset had a large number of features, such as segmentation and wine, the classification accuracy was mainly related to the neighborhood granulation parameter, and the smaller the neighborhood granulation parameter, the worse the classification effect.

In order to compare our algorithms with other classifiers, we ran some experiments on only four datasets. The classification algorithm uses traditional K-Nearest Neighbor (KNN) classifier, and our proposed Neighborhood Granule Classifier was based on Relative (NGCR) distance and the Neighborhood Granule Classifier on Absolute (NGCA) distance. In Table 3, the compared results are shown with a fivefold cross-validation testing for the wine dataset. The test average results of the four UCI datasets are illustrated in Figures 5–8.

Table 3. Classification experiment results for Wine data set.

Testing	KNN	NGCR	NGCA
Accuracy-1	0.9545	0.9545	1
Accuracy-2	0.9636	1	0.9545
Accuracy-3	0.9773	0.9773	1
Accuracy-4	0.9477	0.9545	0.9545
Accuracy-5	0.9294	1	0.9773
Average	0.9545	0.9773	0.9773

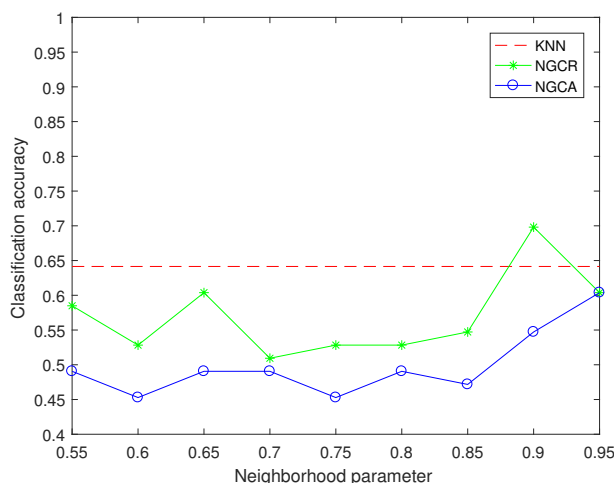


Figure 5. Classification accuracy for the glass dataset.

As can be seen from Figure 5, for the glass dataset, the classification accuracy of KNN was 0.6415, the maximum classification accuracy of NGCR 0.6981, and the maximum classification accuracy of NGCA 0.6038. For the NGCR classifier, the classification accuracy was higher than KNN when the value of neighborhood parameter was 0.9. The NGCR was better than NGCA in any neighborhood parameter value.

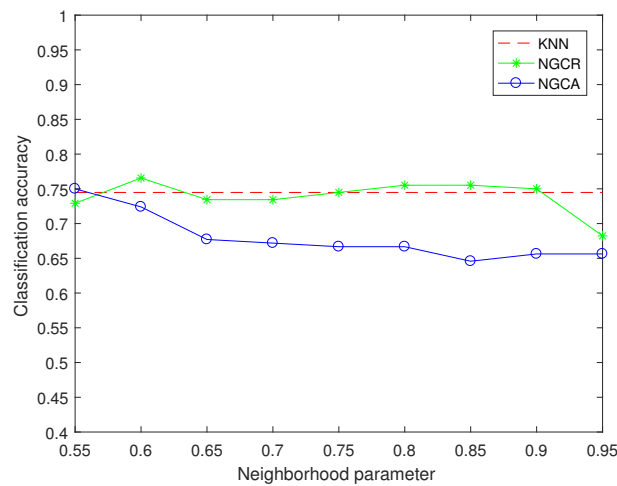


Figure 6. Classification accuracy for the pima dataset.

As can be seen from Figure 6, for the pima dataset, the classification accuracy of KNN was 0.7448. The classification accuracy of NGCR reached a maximum value of 0.7656, when the value of the neighborhood parameter was 0.6; and the classification accuracy of NGCA reached a maximum of 0.75, when the value of the neighborhood parameter was 0.55. Therefore, the NGCR and NGCA granule classifiers were better than the traditional KNN classifier under a suitable neighborhood parameter value.

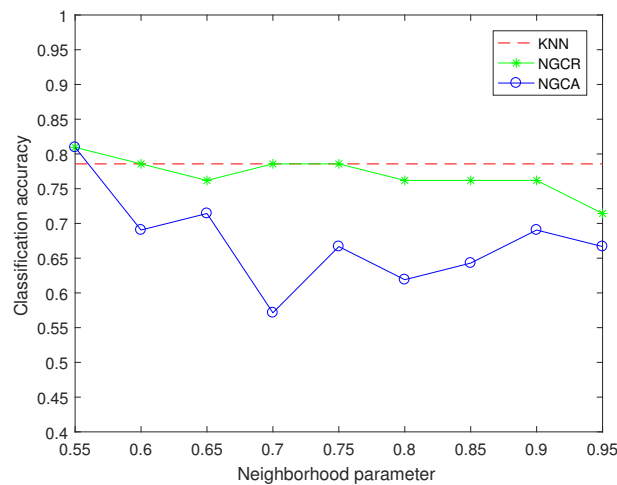


Figure 7. Classification accuracy for the segmentation dataset.

For the segmentation dataset in Figure 7, the classification accuracy of KNN was 0.7857; while the value of the neighborhood parameter was 0.55; the classification accuracy of NGCR reached a maximum value of 0.8095; and that result was the same as NGCA. It can be seen that NGCR and NGCA were better than the traditional KNN under a suitable neighborhood parameter value. The performance of NGCR was better than that of NGCA.

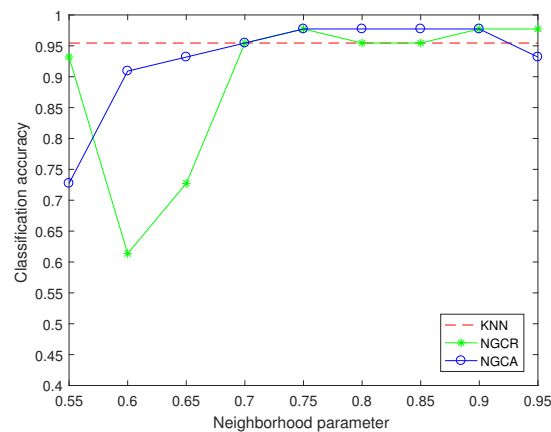


Figure 8. Classification accuracy for the wine dataset.

As for the wine dataset in Figure 8, the classification accuracy of KNN was 0.9545; the classification accuracy of NGCR reached the maximum value of 0.9773, while the neighborhood parameter values were 0.75, 0.90 and 0.95; and while the neighborhood parameter values were 0.75–0.9, the classification accuracy of NGCA reached the maximum value of 0.9773. It can be seen that NGCR and NGCA were better than the traditional KNN under suitable neighborhood parameter values.

6. Conclusions and Discussion

The traditional classifiers are built on the basis of numerical calculations, rather than operations of sets. In this paper, by the neighborhood granulation of samples, a new granule classifier is proposed, which is constructed on the operations of sets. First, the method of neighborhood rough set granulation is introduced to build condition neighborhood granules and decision equivalent granules in decision systems. The size of a granule is defined, the inclusion relationship between the condition granule and the decision granule analyzed, and the granular rule base constructed. Furthermore, we propose two kinds of distance and similarity measures of granules and design an algorithm for a neighborhood granule classifier. Experimental results show that the proposed classifier can successfully classify samples and achieve better classification performance under suitable granulation parameters.

In the future work, a neural network will be introduced to fine-tune the neighborhood granulation parameters that are used for the construction of a classifier. We can also study the granulation of local data to build local neighborhood granules, so as to apply the proposed granule classification method to big data fields.

Author Contributions: Y. C. conceived and designed the experiments; H.J. performed the experiments and wrote the paper.

Funding: This paper is supported by the National Natural Science Foundation of China (61573297), the National Statistical Science Research General Project of China (2018LY64), and the Humanities and Social Sciences Youth Project of Education Ministry of China (18YJC630140).

Acknowledgments: The authors would like to express their thanks to the unknown referees for their careful reading and helpful comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Y.; Wen, K.; Gao, Q.; Gao, X.; Nie, F. SVM based multi-label learning with missing labels for image annotation. *Pattern Recognit.* **2018**, *78*, 307–317. [[CrossRef](#)]
2. Rad, A.B.; Eftestøl, T.; Engan, K.; Irusta, U.; Kvaløy, J.T.; Kramer-Johansen, J.; Wik, L.; Katsaggelos, A.K. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2411–2418. [[CrossRef](#)] [[PubMed](#)]

3. Kim, T.Y.; Oh, K.J.; Sohn, I.; Hwang, C. Usefulness of artificial neural networks for early warning system of economic crisis. *Expert Syst. Appl.* **2004**, *26*, 583–590. [[CrossRef](#)]
4. Tsai, C.F.; Lin, Y.C.; Yen, D.C.; Chen, Y.M. Predicting stock returns by classifier ensembles. *Appl. Soft Comput.* **2011**, *11*, 2452–2459. [[CrossRef](#)]
5. Wang, H.; Gu, J.; Wang, S. An effective intrusion detection framework based on SVM with feature augmentation. *Knowl.-Based Syst.* **2017**, *9*, 1–28. [[CrossRef](#)]
6. Halloran, M.E. Bayes and Empirical Bayes methods for data analysis. *J. Am. Stat. Assoc.* **1997**, *92*, 1640–1651. [[CrossRef](#)]
7. Li, T.; Li, J.; Liu, Z.; Li, P.; Jia, C. Differentially private Naive Bayes learning over multiple data sources. *Inf. Sci.* **2018**, *444*, 89–104. [[CrossRef](#)]
8. Specht, D.F. The general regression neural network-Rediscovered. *Neural Netw.* **1993**, *6*, 1033–1034. [[CrossRef](#)]
9. Bishop, C.M. Bayesian neural networks. *Biol. Cybern.* **2018**, *61*, 361–370. [[CrossRef](#)]
10. Xiao, X.; Jin, L.; Yang, Y.; Yang, W.; Sun, J.; Chang, T. Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition. *Pattern Recognit.* **2017**, *72*, 72–81. [[CrossRef](#)]
11. Bull, L.; Studley, M.; Bagnall, A.; Whittley, I. Learning classifier system ensembles with rule-sharing. *IEEE Trans. Evol. Comput.* **2007**, *11*, 496–502. [[CrossRef](#)]
12. Xu, X.; Zheng, J.; Yang, J.B.; Xu, D.L.; Chen, Y.W. Data classification using evidence reasoning rule. *Knowl.-Based Syst.* **2017**, *116*, 144–151. [[CrossRef](#)]
13. Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **2002**, *21*, 660–674.
14. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: a classification and regression tool for compound classification and qsar modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958. [[CrossRef](#)] [[PubMed](#)]
15. Friedman, N.; Dan, G.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [[CrossRef](#)]
16. Liu, D.; Li, T.; Liang, D. Incorporating logistic regression to decision-theoretic rough sets for classifications. *Int. J. Approx. Reason.* **2014**, *55*, 197–210. [[CrossRef](#)]
17. Zhang, M.L.; Zhou, Z.H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **2007**, *40*, 2038–2048. [[CrossRef](#)]
18. Pang, X.; Xu, C.; Xu, Y. Scaling KNN multi-class twin support vector machine via safe instance reduction. *Knowl.-Based Syst.* **2018**, *148*, 17–30. [[CrossRef](#)]
19. Charniak, E.; Goldman, R.P. A Bayesian model of plan recognition. *Artif. Intell.* **1993**, *64*, 53–79. [[CrossRef](#)]
20. Vapnik, V.; Chapelle, O. Bounds on error expectation for support vector machines. *Neural Comput.* **2000**, *12*, 2013–2036. [[CrossRef](#)]
21. Liu, Y.; Xu, Z.; Li, C. Online semi-supervised support vector machine. *Inf. Sci.* **2018**, *439*, 125–141. [[CrossRef](#)]
22. Miao, X.; Liu, Y.; Zhao, H.; Li, C. Distributed online one-class support vector machine for anomaly detection over networks. *IEEE Trans. Cybern.* **2018**, *99*, 1–14. [[CrossRef](#)] [[PubMed](#)]
23. Shih, P.; Liu, C. Face detection using discriminating feature analysis and support vector machine in video. *Pattern Recognit.* **2006**, *39*, 260–276. [[CrossRef](#)]
24. Cui, P.; Yan, T.T. A SVM-based feature extraction for face recognition. *Pattern Recognit.* **2016**, *43*, 2871–2881.
25. Wang, X.Y.; Wang, T.; Bu, J. Color image segmentation using pixel wise support vector machine classification. *Pattern Recognit.* **2011**, *44*, 777–787. [[CrossRef](#)]
26. Baek, J.; Kim, J.; Kim, E. Fast and efficient pedestrian detection via the cascade implementation of an additive kernel support vector machine. *IEEE Trans. Intell. Transp. Syst.* **2017**, *99*, 1–15. [[CrossRef](#)]
27. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
28. Zhang, L.; Zhou, W.; Wang, B.; Zhang, Z.; Li, F. Applying 1-norm SVM with squared loss to gene selection for cancer classification. *Appl. Intell.* **2017**, *1*, 1–13. [[CrossRef](#)]
29. Hinton, G.E.; Osindero, S.; Teh, Y. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
30. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]

31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 2012 International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
32. Farabet, C.; Couprie, C.; Najman, L.; LeCun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1915–1929. [[CrossRef](#)] [[PubMed](#)]
33. Kavukcuoglu, K.; Sermanet, P.; Boureau, Y.L.; Gregor, K.; Mathieu, M.; Cun, Y.L. Learning convolutional feature hierarchies for visual recognition. In Proceedings of the 2010 International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–11 December 2010; pp. 1090–1098.
34. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
35. Li, H. Deep learning for natural language processing: advantages and challenges. *Natl. Sci. Rev.* **2018**, *5*, 24–26. [[CrossRef](#)]
36. Zadeh, L.A. Fuzzy Logic. *Computer* **1988**, *21*, 83–93. [[CrossRef](#)]
37. Sugeno, M.; Yasukawa, T. A fuzzy-logic-based approach to qualitative modeling. *IEEE Trans. Fuzzy Syst.* **1993**, *1*, 7–31. [[CrossRef](#)]
38. Zadeh, L.A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* **1997**, *90*, 111–127. [[CrossRef](#)]
39. Lin, T.Y. Data mining: Granular computing approach. *Lect. Notes Comput. Sci.* **1999**, *1574*, 24–33.
40. Lin, T.Y.; Zadeh, L.A. Special issue on granular computing. *IEEE Trans. Fuzzy Syst.* **2006**, *14*, 360–360.
41. Yao, Y.Y. Granular Computing using neighborhood systems. In *Advances in Soft Computing*; Springer: London, UK, 1999; Volume 1, pp. 539–553.
42. Yao, Y.Y. Information granulation and rough set approximation. *Int. J. Intell. Syst.* **2001**, *16*, 87–104. [[CrossRef](#)]
43. Yao, Y.; Zhang, N.; Miao, D.; Xu, F. Set-theoretic approaches to granular computing. *Fundam. Inform.* **2012**, *115*, 247–264.
44. Hu, Q.; Yu, D.; Xie, Z. Neighborhood classifiers. *Expert Syst. Appl.* **2008**, *34*, 866–876. [[CrossRef](#)]
45. Hu, Q.; Yu, D.; Liu, J.; Wu, C. Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **2008**, *178*, 3577–3594. [[CrossRef](#)]
46. Zhu, P.; Hu, Q. Adaptive neighborhood granularity selection and combination based on margin distribution optimization. *Inf. Sci.* **2013**, *249*, 1–12. [[CrossRef](#)]
47. Morente-Molinera, J.A.; Mezei, J.; Carlsson, C.; Herrera-Viedma, E. Improving supervised learning classification methods using multigranular linguistic modeling and fuzzy entropy. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1078–1089. [[CrossRef](#)]
48. Pedrycz, W.; Park, B.J.; Oh, S.K. The design of granular classifiers: A study in the synergy of interval calculus and fuzzy sets in pattern recognition. *Pattern Recognit.* **2008**, *41*, 3720–3735. [[CrossRef](#)]
49. Roh, S.B.; Pedrycz, W.; Ahn, T.C. A design of granular fuzzy classifier. *Expert Syst. Appl.* **2014**, *41*, 6786–6795. [[CrossRef](#)]
50. Zhong, C.; Pedrycz, W.; Wang, D.; Li, L.; Li, Z. Granular data imputation: A framework of Granular Computing. *Appl. Soft Comput.* **2016**, *46*, 307–316. [[CrossRef](#)]
51. Pawlak, Z. Rough sets. *Int. J. Inf. Comput. Sci.* **1982**, *11*, 341–356. [[CrossRef](#)]

