IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 26, 2017, accepted March 27, 2018, date of publication April 6, 2018, date of current version April 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2823423

Normalized Cheeger Cut With Neighborhood Rough Approximation

LIN LI AND JIANHUA YUE
China University of Mining and Technology, Xuzhou 221116, China
Corresponding author: Lin Li (lilin@cumt.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 41674133.

ABSTRACT A graph-based collecting study recently attracted significant attention from the scientific research community. Normalized Cheeger cut is a balanced graph partition criterion and a generalized version of normalized graph cut. A stress-free resolution of the normalized Cheeger cut can be obtained by employing the eigenvectors of curve p-Laplacian. However, it is highly sensitive for the original Cheeger cut to collect the interference of noise and disrelated properties. Thus, the performance of the Cheeger cut decreases when high-dimensional data are grouped. To decrease the negative influence of outliers and superfluous properties of collecting, we design an efficient attribute decrease method which is based on neighborhood rough approximation. This design aims to improve the collection of the Cheeger cut. The suggested algorithm introduces information entropy to the neighborhood rough sets in order to measure the importance of attributes. This algorithm reserves the most valuable features and removes the redundant features while retaining the maximum category information of raw data. We then build the p-Laplacian array with the optimized attribute sets and obtain the collecting consequences via the eigen-subspace decomposition of graph p-Laplacian. The cogency of the proposed algorithm is established in various standard data collections. Experimentations demonstrated that our method enjoys sturdy robustness to noise or disrelated feature information in high-dimensional figures.

INDEX TERMS Normalized Cheeger cut, graph p-Laplacian, neighborhood rough set, attribute-value lessening.

I. INTRODUCTION

This era of information explosion constantly exposes us to diverse data every day. Collecting analysis is a powerful tool for data mining and information statistics. The basic idea of collecting is dividing the data set into some categories by determinate analogy measures, which ensures the data belongs to the same collections which own high analogies, whereas the data points that belong to distinguishing collections share low similarities [1]. Statistics collecting aims to explore the core link among objects and find valuable knowledge from massive data. Collecting is the first step in identifying the hidden relationships and models inside data. Collecting has gained increasing attention as a main method of data mining. Collecting in pattern recognition can be used for speech recognition and character recognition; collecting in image processing can be used in image segmentation and machine vision; collecting in statistical learning can be used for data compression and information retrieval [2]. Additionally, collecting has many other applications, such as multi-relational data mining, data stream monitoring, and time series analysis. Collection analysis also plays a significant role in biology, psychology, archaeology, geology, geography, and marketing [3]. Old-style bunching approaches, for example, k-means and FCM algorithm, are appropriate for operating data sets with spherical structures. However, these algorithms fall into local optimum when dealing with non-convex data sets. Graph-based collecting overcomes this limitation by transforming the collecting difficulty to a graph dividing issue. The eigenvectors of curve Laplacian array help obtain a relaxed solution to the graph-cut objective function [4]. At the side of predictable collecting algorithms, graph-based collecting can recognize several complex data structures that are ideal for non-convex data sets. Cheeger cut is a modified form of standardized cut and cheeger cut is gaining increasing consideration [5]. Cheeger cut can produce a high number of well-adjusted collections through graph p-Laplacian array [6]. The p-Laplacian array is not a linear generalization form of Laplacian. Cheeger

cut collecting uses the eigenvectors of p-Laplacian array to group statistics points. Active research is being conducted in this area on account of its reliable theoretical basis and nice collecting consequences. Blekas and Lagaris [7] gained appreciated parallel evidence helping decrease the intersection between categories and amelioration collection eloquence by using Newtons second law to investigate the interface among figures points. Jia *et al*. [8] proposed a fresh scarce spectral collecting technique using the Nystrm approximation and adaptive sampling to reduce the complexity of eigen decomposition. Saade *et al*. [9] used the Bethe Hessian operative to make an improvement of the performance of spectral collecting; they showed that such tactic integrations the merits of the operator which would not track back with tangible symmetric array. Dhanjal *et al*. [10] presented an accessorial spectral collecting that modernizes eigenvectors of Laplacian in a computer-based effective means. Gao *et al*. [11] constructed the scant empathy graph on a trivial typical data collection and used indigenous exclamation to modify the postponement of collecting consequences. Semertzidis *et al*. [12] injected pairwise limitations to a lesser empathy sub-array and used a rare tactic of a milestone phantom collecting to reserve little difficulty. Science and technology have grown significantly and generated massive data that results in data explosionİ. High dimensional is one of the features of these data. It is difficult for customary collecting set of rules to satisfy the demand of present methods in data analysis. Although various collecting algorithms execute in good condition of little figures space, they can not deal with high-dimensional data as well as the low-dimensional data and they often gain unbefitting or invalid collecting consequence [13]. Therefore, it becomes a global research hotspot for people to create and improve neoteric collecting algorithms for excavating immense high-dimension numbers. Attribute reduction is a valid way of reducing data size and it is frequently employed as a pretreatment phase for statistics excavating. Deleting the disrelated or superfluous properties while keep the cataloguing capability of acquaintance base is the substance of the attribute reduction. Attribute reduction reduces computational complicacy and improves the efficiency of algorithms. From the perspective of economy, efficient attribute reduction makes an improvement in acquaintance simplicity of intellectual information structures and decrease the charge of information structures a lot. Such perspective indicates trade principle that cost minimization and profits maximization, which is crucial to commercial cleverness. For the purpose of handling the high-dimensional figures effectively, we design a novel characteristic lessening method relied on neighborhood rough set as well as integrate it with normalized Cheeger cut (NCC) collecting. The anticipated algorithm succeed to the benefits of neighborhood rough approximation and curve p-Laplacian. The comprehensive experiments on target statistics collections prove the effectiveness of this algorithm. This study is arranged as following: Section 2 presents NCC. Section 3 uses neighborhood rough approximation to select the most valuable attributes. Section 4 improves NCC with the optimized attribute set. Section 5 compares the enactment of the suggested algorithm with several prevalent collecting set of rules. Section 6 gives the conclusions of major findings and discusses the direction for future studies.

II. NCC COLLECTING

A. APPLICATION BACKGROUND

Spitting image division is a puzzle in supercomputer visualization. Image dissection is a pivotal step from image processing to image analysis and has a long research history. Image segmentation methods based on collecting analysis are important and widely used image segmentation algorithms. Images in grayscale, color, texture, or other types can be segmented by collecting method [14]. The main idea of image segmentation is performing collecting on the image pixels to obtain the segmentation consequence. Initially, the pixels in image space is represented with the corresponding points of feature space. After the collecting in feature space, we map the collecting consequences returning to the primary image room. Researchers have presented different interpretations and expressions of image segmentation. Using the concept of setting, image segmentation can be regularly defined [15]. Make set R represent the entire image area. The dissection of R could be regarded as the process of allocating R into k non-empty sets (sub-regions) R1, R2,..., Rk which satisfy the following five circumstances.

(1) $\cup_{i=1}^{n} Ri = R$

(2) For every$i \neq j$, there is

(3) For i = 1, 2, k, there is $\mathrm{P}(R_i) = \mathrm{TRUE}$

(4) For $i \neq j$, there is $\mathrm{P}(R_i \cap R_j) = \mathrm{FALSE}$

(5) For i = 1, 2, k, R_i is a connected region

where $\mathrm{P}(R_i)$ is the logical predicate for all the elements in set R_i, and $\emptyset$ is the empty set. Condition (1) means the sum (union) of all sub regions in the image segmentation consequences should include all pixels in the original image. That is, segmentation should ensure that every pixel in the appearance is grouped into a sub-area. Condition (2) indicates that each subregion in the segmentation consequences does not overlap each other. A pixel cannot belong to two regions. Condition (3) means that each subregion in the segmentation consequences has unique characteristics or that pixels in the same region have the same characteristics. Condition (4) indicates the lack of public cross parts among regions in the segmentation consequences. Condition (5) shows that the pixels in the same sub region of the segmentation consequences should be connected. The above definitions and the definition of collecting have similarities. Thus, image segmentation problem can be identified as a collecting problem. We may collection on the pixel feature set of image to obtain the final segmentation consequence. Collecting analysis can address the lack of training samples in image segmentation task and meet the conditions of unsupervised processing requirements [16]. Image segmentation method based on collecting usually consists

of two steps: (1) the feature extraction process of pixels and (2) treating image pixels as sample points and applying the collecting algorithm to divide them into collections. Cheeger cut collecting has a solid theoretical foundation and can identify complex data structures. This method also has great advantage in image segmentation.

B. OBJECTIVE FUNCTION OF CHEEGER CUT

Cheeger cut collecting is based on ethereal graph model. Owned a statistics set, we structure an directionless slanted curve G = (V,E), where V is the set of vertices signified by statistics dots, and E is the regular of edges weighted by the parallels in the middle of the crossed vertices in the edge. Assume A is a subset of V, the counterpart of A is inscribed as $\bar{A} = V \backslash A$ The censored of A and $\bar{A}$ is well-demarcated as

$$cut(A, \overline{A}) = \sum_{i \in A, j \in \overline{A}} W_{ij} \tag{1}$$

where wij is the resemblance between vertices i and j. A clear graph part represents that the similarities within a gathering are the largest, whereas the similarities among collections are the smallest. To obtain balanced collections, Bühler and Hein [17] modified the normalized cut criterion and propose normalized Cheeger cut, which is denoted as NCC.

$$NCC(A, \overline{A}) = \frac{cut(A, \overline{A})}{min\{vol(A), vol(\overline{A})\}} \tag{2}$$

where $vol(A) = \sum_{i \in A, j \in V} w_{ij}$ is the volume of subset A. Formula (2) considers the internal and external connections of collections. A Cheeger cut divides the graph into a number of subgraphs to minimize Formula (2). However, research indicates that it is an NP-hard problem to seek the optimal solution of a normalized Cheeger cut. Through bringing in a p-Laplacian array, we may obtain a hassle-free resolution of normalized Cheeger cut, which is stated by Rayleigh quotient principle.

C. GRAPH P-LAPLACIAN

Spectral curve model is closely connected with the graph Laplacian array. We can utilize the properties of the p-Laplacian array to optimize the objective function of a NCC. It is defined by Hein *et al*. [18] that the internal creation formula of customary graph Laplacian Δ_2 as follows.

$$\langle f, \bigtriangleup_2 f \rangle = \frac{1}{2} \sum_{i,j=1}^{n} W_{ij} (f_i - f_j)^2 \tag{3}$$

where f is the eigenvector of Laplacian matrix. Assuming the Laplacian operator is generalized to Δ_p, where $p \in (1, 2]$, then Δ_p is denoted as

$$\langle f, \bigtriangleup_p f \rangle = \frac{1}{2} \sum_{i,j=1}^{n} W_i j (f_i - f_j)^p \tag{4}$$

The normalized graph Laplacian Δ_2 in matrix notation is represented as $\Delta_2 = I - D^{-1}W$, where W is the weight matrix formed by edge weights. D is the degree matrix, which is a diagonal matrix with the diagonal element $d_i = sum_{i=1}^{n} W_{ij}$. The corresponding normalized p-Laplacian operator Δ_p can be derived easily:

$$(\bigtriangleup_p f)_i = \frac{1}{d} \sum_{j \in V} W_{ij} \varphi_p (f_i - f_j) \tag{5}$$

where $\varphi_p(y) = |y|^{p-1} sign(y)$, $y \in r$, and $\varphi_2(y) = y$ when p = 2. The eigenvalue λ_p of p-Laplacian is defined as follows.

Definition 1: If there is a real number λ_p that satisfies Formula (6), λ_p is the corresponding eigenvalue of Eigenvector f.

$$(\bigtriangleup_p f)_i = \lambda_p \phi_p (f_i), \quad \forall i = 1, \ldots, n \tag{6}$$

In a array operation, the eigenvector that corresponds to the minutest eigenvalue is asked to include significant discernment information. This attribute is crucial in a collecting algorithm as well. Definition 2 applies this characteristic to a nonlinear p-Laplacian operator.

Definition 2: If λ_p is the smallest eigenvalue of normalized graph p-Laplacian Δ_p, λ_p should satisfy Formula (7) to reach the lower bound.

$$\lambda_p = \arg\min_{f \in R} \frac{\langle f, \Delta_p f \rangle}{\sum_{i=1}^{n} d_i |f_i|^p} \tag{7}$$

Formula (7) can be converted into a function Gp(f) of normalized p-Laplacian as follows.

$$G_p(f) = \frac{\langle f, \bigtriangleup_p f \rangle}{\sum_{i=1}^{n} d_i |f_i|^p} = \frac{\frac{1}{2} \sum_{i,j=1}^{n} W_{ij} |f_i - f_j|^p}{\sum_{i=1}^{n} d_i |fi|^p} \tag{8}$$

The analysis from above proves that the inferior bound of Gp(f) is relative to the eigenvalues and eigenvectors of Δ_p. We then show the mathematical connection between Gp(f) and the objective function of NCC.

D. RELAXED SOLUTION OF NCC

Theorem 1: For p>1 and every partition of V into A, $\bar{A}$, a function (f,A) exists such that the functional Gp associated to the normalized p-Laplacian satisfies

$$G_p(f, A) = cut(A, \overline{A}) \left| \frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\overline{A})^{\frac{1}{p-1}}} \right|^{p-1} \tag{9}$$

Formula (11) can be understood as a harmonious graph cut principle. We then obtain the following special case:

$$\lim_{p \to 1} G_p(f, A) = NCC(A, \overline{A}) \tag{10}$$

Proof: Function (f,A) for a partition A, of V is first defined as

$$(f, A)_i = \begin{cases} 1/vol(A)^{\frac{1}{p-1}} & i \in A \\ -1/vol(\bar{A})^{\frac{1}{p-1}} & i \in \bar{A} \end{cases} \tag{11}$$

Formula (11) is brought into $\langle f, \Delta_p f \rangle$ and $\sum_{i=1}^{n} d_i |fi|^p$. Thus, we have

$$\begin{aligned}
\langle f, \Delta_p f \rangle &= \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} |f_i - f_j|^p \\
&= \sum_{i \in A, j \in \bar{A}} w_{ij} \left| \frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}} \right|^p \\
\sum_{i=1}^{n} d_i |f_i|^p &= \sum_{i \in A} d_i \left| \frac{1}{vol(A)^{\frac{1}{p-1}}} \right|^p + \sum_{i \in \bar{A}} d_i \left| \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}} \right|^p \\
&= \frac{vol(A)}{vol(A)^{\frac{p}{p-1}}} + \frac{vol(\bar{A})}{vol(\bar{A})^{\frac{p}{p-1}}} \\
&= \frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}}
\end{aligned}$$

By substituting the numerator and denominator of Formula (8) with the above two expressions, we acquire the formula:

$$\begin{aligned}
G_p(f, A) &= \frac{\langle f, \Delta_p f \rangle}{\sum_{i=1}^{n} d_i |f_i|^p} \\
&= \frac{\sum_{i \in A, j \in \bar{A}} w_{ij} \left| \frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}} \right|^p}{\frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}}} \\
&= \sum_{i \in A, j \in \bar{A}} w_{ij} \left| \frac{1}{vol(A)^{\frac{1}{p-1}}} + \frac{1}{vol(\bar{A})^{\frac{1}{p-1}}} \right|^{p-1} \\
&\leq \sum_{i \in A, j \in \bar{A}} w_{ij} \left| \frac{2}{\min\{vol(A), vol(\bar{A})\}^{\frac{1}{p-1}}} \right|^{p-1} \\
&= 2^{p-1} \frac{cut(A, \bar{A})}{\min\{vol(A), vol(\bar{A})\}}
\end{aligned}$$

Make a comparison between the above inequality and the objective function of NCC obtains

$$\lim_{p \to 1} G_p(f, A) = NCC(A, \bar{A})$$

End Proof

Theorem 1 shows that the solution of Gp(f) is a relaxed estimated resolution of $NCC(A, \bar{A})$. From the process of minimizing Gp(f). the optimal graph partition can be gained. The extreme value of Gp(f) is the eigenvalue of graph p-Laplacian.

$$\lambda_p = \arg\min_{p \to 1} G_p(f) \tag{12}$$

where λ_p is the eigenvalue corresponding to eigenvector f.

Thus, using the p-Laplacian operator, NCC can be figured out in polynomial time. After set a suitable threshold, the second eigenvector ${v_p}^{(2)}$ of p-Laplacian array leads to a bipartition of the graph [17]. After minimized the corresponding Cheeger cut, the optimal threshold is decided. The following threshold should satisfy the second eigenvector ${v_p}^{(2)}$ of graph p-Laplacian Δ_p:

$$\arg\min_{A_t = \{i \in V | v_p^{(2)}(i) > t\}} NCC(A_t, \bar{A}_t) \tag{13}$$

III. NEIGHBORHOOD ROUGH SET APPROXIMATION

Pawlak [19] proposed rough set theory in 1982. Conceptions of domain, lower estimate and upper calculation are included in the definition of the theory in order to designate the progression of anthropological book learning and critical reasoning. In a rough set model, production rules represent obtained knowledge and it is simple for operators to comprehend, receive and employ. For choosing property subsets, searching decision regulations and exploring acquaintance reliance and other arenas, rough set is used generally as a calculated instrument for handling blurry and indeterminate acquaintance. Among the rough set knowledge discovery, attribute reduction is the main content. This process describes the necessity of each attributes in the information system and the process of removing superfluous knowledge.

Nevertheless, the foundation of the Pawlak rough set [19] is known as the characteristic philosophy of uniformity associations and similarity categories which are merely appropriate for dealing with discrete data. Constant figures are general in the real domain and they should be discretized at the early stage. We could choose suitable dissection in order to separate the series of nonstop property principles into a variety of separation breaks and then show the characteristic values in every subinterval by using different integers. The loss of information inevitably happened in such conversion and the analyzing consequences are based on the efficacy of the separation intervals to a huge extent. In order to figure out this puzzle, Hu *et al*. [20] introduced neighborhood relationships into a rough set and proposed the neighborhood rough set exemplary. Attributes can be analyzed straightly in this model with unbroken values to eliminate the discretization procedure. Consequently, there are a lot of benefits in feature assortment and cataloguing accuracy.

A. δ-NEIGHBORHOOD

Rough sets define the difficult that will be dealt with as information organism. $S = \langle U, A, V, F \rangle$ is an information system. U is domain, which is a non-empty data set. A is the attribute set. V is the feature value set. F is a map function, which indicates the association of the model and the relevant characteristic.

Definition 3: Jia, Ding, Ma and Xing(2014) point out that given domain U, for object $y_i \in U$, the definition of δ-neighborhood of y_i is

$$\delta(y_i) = \{y | y \in U, \Delta(y, y_i) <= \delta\} \tag{14}$$

where $\delta \geq 0$, $\delta(y)$ is the neighborhood particle of y_i, and Δ is a distance function. For $y_1, y_2, y_3 \in U$, Δ satisfies the following relations.

(1) $\Delta(y_1, y_2) \geq 0$, $\Delta(y_1, y_2) = 0$,, if and only if $y_1 = y_2$;
(2) $\Delta(y_1, y_2) = \Delta(y_2, y_1)$;
(3) $\Delta(y_1, y_3) \leq \Delta(y_1, y_2) + \Delta(y_2, y_3)$. [21]

For a sample set of N attributes, distance is usually calculated by P-norm.

$$\Delta_p(y_1, y_2) = (\sum_{i=1}^{n} |f(y_1, a_i) f(y_2, a_i)|^p)^{\frac{1}{p}} \tag{15}$$

where f(y, a_i) is the value of attributes ai of sample y. If a_i is a symbolic trait,

(1) $|f(y_1, a_i) - f(y_2, a_i)| = 0$, if y1, y2 have the same value on attribute a_i; and

(2) $|f(y_1, a_i) - f(y_2, a_i)| = 1$, if y1, y2 have disparate values on attribute a_i.

For instance, domain $U = \{y_1, y_2, y_3, y_4, y_5\}$, a is an attribute of U and $f(y, a)$ stands for the attribute value of sample y on attribute a. $f(y_1, a) = 1.1$, $f(y_2, a) = 1.2$, $f(y_3, a) = 1.6$, $f(y_4, a) = 1.8$, and $f(y_5, a) = 1.9$. Neighborhood size is set to $\delta = 0.2$. Given that $|f(y_1, a) - f(y_2, a)| \leq 0.2$;thus, $y_2 \in \delta(y_1)$, $y_1 \in \delta(y_2)$. We then derive each samples δ-neighborhood: $\delta(y_1) = \{y_1, y_2\}$, $\delta(y_2) = \{y_1, y_2\}$, $\delta(y_3) = \{y_3, y_4\}$, $\delta(y_4) = \{y_3, y_4, y_5\}$, $\delta(y_5) = \{y_4, y_5\}$. If domain U includes multiple attributes, the δ-neighborhood of samples can be calculated in a similar way.

B. NEIGHBORHOOD DECISION SYSTEM

Definition 4: Given a domain $U = \{y_1, y_2, \cdots\cdots, y_n\}$ positioned in real space, A is on behalf of the attribute set of U and D is on behalf of the decision-making attribute. If A can produce a type of neighborhood connection of domain U, then $NDT = \langle U, A, D\rangle$ is named a neighborhood decision system. For a neighborhood decision system $NDT = \langle U, A, D\rangle$, eparated to N similarity categories by decision-making attribute D:$y_1, y_2, \cdots\cdots, y_n$. $\forall B \subseteq A$, the upper approximation,lower approximation, and decision-making borderline of decision attribute D about B are correspondingly distinct as

$$\underline{N_B}D = \bigcup_{i=1}^{N} \underline{N_B} Y_i \tag{16}$$

$$\overline{N_B}D = \bigcup_{i=1}^{N} \underline{N_B} Y_i \tag{17}$$

$$\overline{N_B}D = \bigcup_{i=1}^{N} \overline{N_B} Y_i \tag{18}$$

where $\overline{N_B}Y_i = \{y_i | \delta_B(y_i) \cap Y_i \neq \emptyset, y_i \in U\}$, and $\underline{N_B}Y_i = \{y_i | \delta_B(y_i) \subseteq Y_i, y_i \in U\}$. The concept of upper and lower approximation is explained by giving a case of the categorizations of two types. Assume that domain U includes two correspondence types, just as seen in Figure 1. One type of testers is labelled with *, the other are labelled with +. The figure shows that the tasters in the rounded neighborhood of taster y_1 are belong to category *. Thus, y_1 is part of the lower approximation of class *. The tasters in the neighborhood of y_3 come from category +. Thus, y_3 fits in the lower guesstimate of category +. Category * tasters and category + samples are covered in the neighborhood of sample y_2. Thus, y_2 is a borderline taster. This categorization is accordant with our instinctive understanding of categorization difficulties.

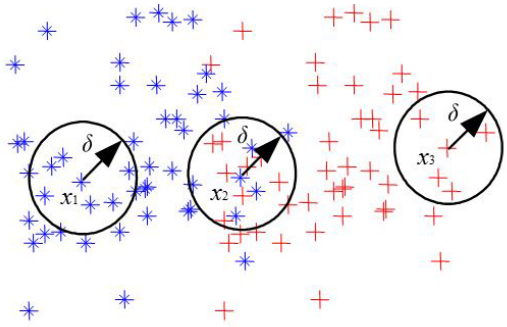

FIGURE 1. Neighborhood rough set exemplary.

The lower approximation $\underline{N_B}D$ of decision attribute D is also called a positive decision region denoted as $POS_B(D)$. The size of $POS_B(D)$ mirrors the separable point of domain U in a given attribute space. A large positive region consequence in sharp boundaries of each category and less overlap.

We describe the reliance of decision-making attribute D on condition attribute B on the basis of the nature of the positive region.

$$\gamma_B(D) = \frac{Card(\underline{N_B}D)}{Card(U)} \tag{19}$$

where $0 \leq \gamma_b(D) \leq 1$. $\gamma_b(D)$ is the proportion which is obtained from the tasters entirely confined in a definite category of decision-making accounted for every tester, on the basis of the description of situation attribute B, in the tester series. The superior the positive region $\underline{N_B}D$ is, the reliance of decision D on circumstance B will be stronger.

C. NEIGHBORHOOD ATTRIBUTE REDUCTION

Jia *et al*. claim that given $NDT = \langle U, A, D\rangle$, B is a subset of A. For an arbitrary attribute, $\forall a \in B$, If $\gamma_{B-a}(D) < \gamma_B(D)$, a is indispensable to B. If $\gamma_{B-a}(D) = \gamma_B(D)$, a is an irrelevant feature. [21]If the attributes in B are all indispensable to B, B is a reduced set.

Definition 5: Jia *et al*. claim that given a $NDT = \langle U, A, D\rangle$, if B is a subset of A and it meets the conditions below, B is called A's reduction. (1) $\forall a \in B$, $\gamma_{B-a}(D) < \gamma_B(D)$, and (2) $\gamma_A(D) = \gamma_B(D)$. [21] Condition (1) means that there are no irrelevant attributes in a reduction set, namely the reduction set should be isolated. Condition (2) indicates that the reduction process should not influence the whole distinguishable nature of the system.

Definition 6: Given a neighborhood decision system $NDT = \langle U, A, D\rangle$, $B \sqsubseteq A$, and $\forall a \in A - B$, the important degree of a relative to B is defined as

$$SIG(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D) \tag{20}$$

According to the neighborhood rough sets, we can create the attribute decrease algorithm by using the traits significance index. First, we work out the significance degree of the rest of traits and enlarge the attribute with the most significance to the decrease series. This practice is repeated until the significant degree of the rest of attributes is 0, which indicates that the dependent function values of the system will not change when a new attribute is added. However, some attributes may own the highest degree of significance. Traditional reduction algorithms choose one attribute from all attributes randomly, which arbitrarily does not consider the influences on additional elements on attribute assortment and might consequence in bad discount consequences.

Based on information theory, analyzing attribute reduction make an improvement on reduction accuracy. Several researchers have proven the feasibility of this method. Wu and Gou [22] proposed a decision table reduction algorithm based on conditional information entropy. Shannon [23] introduced joint information into a decision table to designate attribute significance and gain the correlative decrease of attributes. Shannon, as the inventor of information concept, points out that any information has severance [24]. There is a close relationship among the size of the redundancy and the incidence likelihood, improbability of each emblem of information, such as statistics, letters and characters. Information eliminates the indecision of possessions. The higher the vagueness is, the amount of the entropy value will be greater and the amount of information needed to achieve clarity will be larger [25]. The characterization of entropy is given below.

Definition 7: Given knowledge P and its partition $U/P = \{Y_1, Y_2, \cdots\cdots, Y_n\}$ exported on domain U, the information entropy of knowledge P is defined as:

$$H(P) = -\sum_{i=1}^{n} p(Y_i) log p(Y_i) \tag{21}$$

where $P(Y_i) = |Y_i|/|U|$ is on behalf of the probability of equivalence category Y_i on domain U. We appraise attributes by using information entropy which is considered as another criterion. When numerous attributes have the highest degree of significance, we make a comparison between the info entropy of such attributes and choose the attribute with minutest entropy which means that bringing the minimum indeterminate information, then incorporate it into the lessening set to obtain improved attribute lessening consequences. This modified attribute lessening algorithm is exposed as Algorithm 1.

Algorithm 1: Attribute granulation based on neighborhood rough set approximation. Input: the $NDT = \langle U, A, D\rangle$ system.

Output: attribute reduction result *red*.

Step 1. For each attribute a in attribute set A, compute its relationship N_a with neighborhood.

Step 2. Use empty set to initialize *red*.

Step 3. Set $B = A - red$. For each attribute $a \in B$, computes its importance degree $SIG(a_i, red, D) = \gamma_{red \cup a}(D) - \gamma_{red}(D)$.

Step 4. If there is only one attribute a_k has the maximum importance degree, namely $SIG(a_k, red, D) = max_i(SIG(a_i, red, D))$, a_k is a candidate attribute; otherwise, we need to compare the entropy of each attributes and select the attributes that satisfy $H(a_k) = min_i(H(a_i))$ (Jia *et al*., 2014). [21]

Step 5. If the importance degree of a_k is greater than 0, add a_k to the set *red*, then turn to Step 3; otherwise, output the reduction set *red*(Jia *et al*., 2014). [21]

IV. NCC WITH NEIGHBORHOOD ROUGH APPROXIMATION

Huge high-dimensional statistics handing out has been a puzzle in statistics excavating. Statistics of high-dimension is often supplemented by the obscenity of dimensionalityİ. Thus, old NCC cannot fully play its advantages. Furthermore, there are a mass of noise and disrelated features in the real data sets, which probably causes dimension trapİ. The dimension trapİ would impede with the collecting procedure of algorithms and influence the accuracy of collecting consequences [26]. Thus, we propose a novel NCC algorithm with neighborhood rough approximation (NRA-NCC).This algorithm first computes the degree of importance of each attribute based on neighborhood rough set theory. This algorithm then removes redundant attributes, retain the most important attributes according to their importance and the characteristics of the decision system, measure the resemblances among statistics points to structure the likeness array and p-Laplacian array, and divide the graph into multiple subgraphs with the eigenvectors of p-Laplacian array to minimize the normalized Cheeger cut specification and acquire great superiority collecting consequences. The thorough phases of NRA-NCC algorithm are shown in Algorithm 2.

Algorithm 2. NCC with neighborhood rough approximation

Input: Dataset, the cluster number k

Output: k separated clusters

Step 1. Decrease the attributes of data points according to Algorithm 1 and gain the reduced attribute set red.

Step 2. After attribute granulation, compute the resemblances among data points according to the new data set red and form the sympathy array $W \in R^{n\times n}$ using self-tuning Gaussian kernel function, which is defined by the following formulation:

$$W_{ij} = exp\left(-\frac{d^2(y_i, y_j)}{\sigma_i \sigma_j}\right) \tag{22}$$

where $\sigma_i = \frac{1}{p}\sum_{k=1}^{p} d(y_i, y_k)$ is the Euclidean distance from point y_i to its p-th nearest neighbor. Rather than use a unchanging parameter σ, this algorithm sets a homologous stricture σ_i for each point y_i which is based on their neighborhood information.

Step 3. Modify the first collection $A_i = V$ and set the collection number s = 1.

Step 4. Repeat Steps 4 to 8.

TABLE 1. Data sets used in the experiments.

Data set	Instance number	Condition attribute number	Decision attribute number	Class number
Ionosphere	351	34	1	2
Sonar	208	60	1	2
WDBC	569	30	1	2
Colon Cancer	62	2000	1	2
Duke Breast Cancer	44	7129	1	2
Leukemia	72	7129	1	2

Step 5. Based on Formula (5) with the affinity array W to structure p-Laplacian array.

Step 6. Figure out the second eigenvector $V_p^{(2)}$ of graph p-Laplacian Δ_p and seek a suitable threshold value which meets Formula (13).

Step 7. Use $V_p^{(2)}$ to split every collection $A_i(i = 1, 2 \cdots\cdots, S)$ and minimalize the general Cheeger cut objective function.

Step 8. $s \Leftarrow s + 1$.

Step 9. Terminate the circle and output the collecting consequences until the number of collections $s = k$.

V. EXPERIMENTAL ANALYSIS

A. DATA SETS

We take six benchmark statistics series to conduct the trials and test the effectiveness of the proposed NRA-NCC algorithm. The traits of such statistics series are seen in Table 1. Ionosphere, Sonar, and WDBC are from the University of California Irvine machine learning repository[1]. Colon Cancer, Duke Breast Cancer, and Leukemia are all cancer data sets from the LIBSVM data page [2]

B. EVALUATION METRIC

Several methods can measure the merits of collecting consequences. F-measure is a commonly used evaluation index which is on the basis of artificial annotation. F-measure is acquired from information repossession arena and includes accuracy and recollection ratios. The distinction between the collecting consequences and the realistic categories from various angles are described by these two indicators. F-score is figured out by the accuracy and recollection ratio is a widespread index to assess a collection. F-score gives an unbiased assessment to the engendered collections. Assume that k categories are in the dataset, and category I is connected with collection i* in collecting consequences. We may calculate the F-score of category i by executing the following three formularies.

$$P(i) = N_{ii^*}/N_{i^*} \tag{23}$$

$$R(i) = N_{ii^*}/N_{i^*}, \tag{24}$$

and

$$F(i) = \frac{2 \times P(i) \times R(i)}{P(i) + R(i)} \tag{25}$$

where P(i) and R(i) are separately the accuracy ratio and the recollection ratio, N_{ii^*} represents the scope of the juncture of category i and collection i*, Ni is the size of category i, and Ni* is the magnitude of collection i*.

The complete F index of the collecting consequences is the weighted regular of every categorys F-score.

$$F = \frac{1}{n} \sum_{i=1}^{k} [N_i \times F(i)] \tag{26}$$

where n is the number of taster points, k is the category number of statistics set, and Ni is the proportions of category i. $F \in [0, 1]$. With the growth of the F index, the gap between the collecting consequences of the algorithm and the real data category will be narrowed.

C. COLLECTING CONSEQUENCES

In the trial, NRA-NCC algorithm is compared with the formal normalized cut collecting (NC) density adaptive spectral collecting (DSC) [27], self-tuning p-spectral collecting (ST-pSC) [28] and NCC algorithm [5]. Figure 2 shows the collecting consequences of collecting consequences of five algorithms on each data set. The horizontal axis of the figure is the collection label and the vertical axis is the F-score of every collection.

Figure 2 shows that the enactment of NC algorithm is near NCC algorithm. NC and NCC are on the basis of graph theory and collecting problematic is transformed into a graph segregating difficult. So these two algorithms might discover the international optimal resolution with the assistance of the spectral technique and Laplace transform.DSC algorithm uses local density adaptive similarity measure to calculate the similarities between data points. ST-pSC algorithm classifies data points based on shared nearest neighbors. NCC algorithm can generate balanced collections on Ionosphere data set. DSC algorithm is suit for Sonar data set. ST-pSC algorithm works well on WDBC dataset. However, the F-scores of NC, DSC, ST-pSC and NCC for high dimensional collecting problems are lower than the proposed NRA-NCC algorithm. The information in every attribute of the cases is distinguishing and they play different

[1] http://archive.ics.uci.edu/ml/

[2] http://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets/

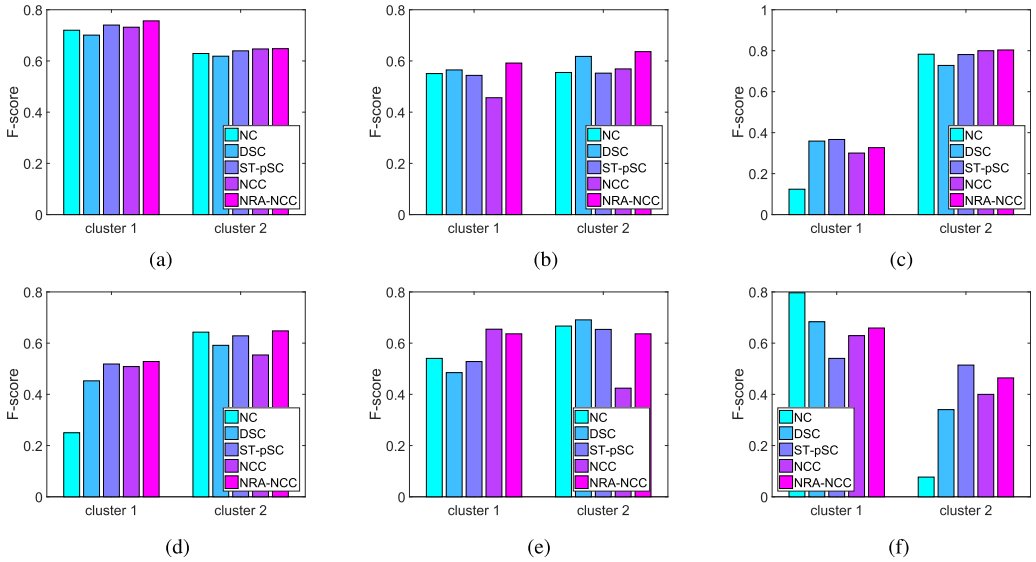

FIGURE 2. Collecting consequences on different datasets. (a) Ionosphere. (b) Sonar. (c) WDBC. (d) Colon Cancer. (e) Duke Breast Cancer. (f) Leukemia.

TABLE 2. Total F index of different algorithms.

Data set		Algorithm				
		NC	DSC	ST-pSC	NCC	NRA-NCC
Ionosphere	F index	0.6874	0.6390	0.7040	0.7013	0.7177
	Condition attributes	34	34	34	34	15
Sonar	F index	0.5530	0.5622	0.5483	0.5165	0.6156
	Condition attributes	60	60	60	60	23
WDBC	F index	0.5374	0.5608	0.6593	0.6139	0.6260
	Condition attributes	30	30	30	30	12
Colon Cancer	F index	0.5035	0.5423	0.5895	0.5362	0.6055
	Condition attributes	2000	2000	2000	2000	136
Duke Breast Cancer	F index	0.6007	0.5832	0.5901	0.5446	0.6364
	Condition attributes	7129	7129	7192	7192	201
Duke Breast Cancer	F index	0.5467	0.5753	0.5314	0.5496	0.5915
	Condition attributes	7129	7129	7129	7129	248

roles in the collecting. Improper feature selection will largely affect the collecting consequences. Traditional graph cut collecting algorithm does not take this into account. This algorithm is impressionable to the influence of noise and disrelated attributes and are not suitable for high-dimensional applications. However, with the help of attribute granulation, NRA-NCC algorithm can handle multi-scale collecting problems and identify more complex data structures. For additional evaluation, Table 2 shows the whole F index for every algorithm and the amount of condition attributes of distinguishing data sets.

Table 2 shows that the NRA-NCC algorithm works well with high-dimensional data. On most data sets, the proposed NRA-NCC algorithm can generate more accurate collecting consequences than NC, DSC, ST-pSC and NCC algorithms. Neighborhood rough sets are used by NRA-NCC algorithm to adjust data occasions. Neighborhood attribute granulation deletes attributes which have no relation and keeps the attributes that make the most contribution to collecting. Thus, the statistics points within the equal collection are compressed, whereas the statistics points between distinctive collections are divided. Thus, NRA-NCC algorithm has high collecting accuracy in most cases. The neighborhood attribute reduction is on the basis of information entropy reduces the passive influences of noise figures and superfluous attributes on the collecting. This method reduces the difficulty of problem solving and describes the approximate relationship between data points much better. NRA-NCC algorithm integrates the advantages of Cheeger cut collecting and neighborhood rough approximation. This algorithm has good robustness and strong generalization ability.

VI. CONCLUSIONS

Discrete data is the only appropriate date to apply outmoded rough set philosophy, but old rough set theory requires the discretization of the domain when dealing with continuous

data. Neighborhood rough sets overcome this limitation through neighborhood approximation, which can directly process the numeric statistics. We adjust the attribute lessening technique which is on the basis of neighborhood rough sets to improve the performance of NCC on high-dimensional data. Selecting the suitable attributes with the new method, there is a close connection between the attribute importance and the information entropy. Facing multiple attributes with the equal significance degree, we may make a comparison between the information entropy of these traits and choose the attribute with trifling entropy to join the lessening set to improve the compact attribute set. We propose that the NRA-NCC algorithm is rely on the optimized attribute reduction set in the second innovation. NRA-NCC highlights the distinctions between data points while keeping their characteristics to let the last collecting consequences close to the realistic statistics categorization. Experimentations prove that the proposed NRA-NCC algorithm is grander to traditional Cheeger cut algorithms. NRA-NCC has strong anti-jamming and good generalization abilities to high-dimensional data. Future study will focus on how to apply NRA-NCC algorithm to a recommendation system, text categorization, pattern recognition, and other fields.

AUTHOR CONTRIBUTIONS STATEMENT

Li Lin: algorithm design, experiments, paper writing. MiJia Wang: experiments, paper writing. Yue Jianhua: algorithm design, data analysis, paper writing.

ADDITIONAL INFORMATION

Competing financial interests: The authors declare no competing financial interests.

REFERENCES

[1] H. Jia, S. Ding, X. Xu, and R. Nie, "The latest research progress on spectral clustering," *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1477–1486, 2014.
[2] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1356–1363.
[3] M. Liu, H. Liu, Q. Sun, T. Zhang, and R. Ding, "Salient pairwise spatio-temporal interest points for real-time activity recognition," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 14–29, 2016.
[4] H. Jia, S. Ding, M. Du, and Y. Xue, "Approximate normalized cuts without Eigen-decomposition," *Inf. Sci.*, vol. 374, pp. 135–150, Dec. 2016.
[5] X. Bresson and A. Szlam, "Total variation and Cheeger cuts," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 1039–1046.
[6] X. Bresson, T. Laurent, D. Uminsky, and J. H. von Brecht, "Multiclass total variation clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1421–1429.
[7] K. Blekas and I. E. Lagaris, "A spectral clustering approach based on Newton's equations of motion," *Int. J. Intell. Syst.*, vol. 28, no. 4, pp. 394–410, 2013.
[8] H. Jia, S. Ding, and M. Du, "A Nyström spectral clustering algorithm based on probability incremental sampling," *Soft Comput.*, vol. 21, no. 19, pp. 5815–5827, 2017.
[9] A. Saade, F. Krzakala, and L. Zdeborová, "Spectral clustering of graphs with the Bethe Hessian," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 406–414.
[10] C. Dhanjal, R. Gaudel, and S. Clémençon, "Efficient Eigen-updating for spectral graph clustering," *Neurocomputing*, vol. 131, pp. 440–452, May 2014.
[11] J. Cao, P. Chen, Q. Dai, and W.-K. Ling, "Local information-based fast approximate spectral clustering," *Pattern Recognit. Lett.*, vol. 38, pp. 63–69, Mar. 2014.
[12] T. Semertzidis, D. Rafailidis, M. G. Strintzis, and P. Daras, "Large-scale spectral clustering based on pairwise constraints," *Inf. Process. Manage.*, vol. 51, no. 5, pp. 616–624, 2015.
[13] L.-A. Gottlieb and R. Krauthgamer, "A nonlinear approach to dimension reduction," *Discrete Comput. Geometry*, vol. 54, no. 2, pp. 291–315, 2015.
[14] X. Wang, Y. Tang, S. Masnou, and L. Chen, "A global/local affinity graph for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1399–1411, Apr. 2015.
[15] C. Tang, R. Wang, Z. Li, W. Wang, and W. Gao, "Frame interpolation with pixel-level motion vector field and mesh based hole filling," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 72–78, 2016.
[16] H. Jia, S. Ding, L. Meng, and S. Fan, "A density-adaptive affinity propagation clustering algorithm based on spectral dimension reduction," *Neural Comput. Appl.*, vol. 25, nos. 7–8, pp. 1557–1567, 2014.
[17] T. Bühler and M. Hein, "Spectral clustering based on the graph p-Laplacian," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 81–88.
[18] M. Hein, J.-Y. Audibert, and U. V. Luxburg, "Graph Laplacians and their convergence on random neighborhood graphs," *J. Mach. Learn. Res.*, vol. 8, pp. 1325–1368, Jun. 2007.
[19] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[20] Q.-H. Hu, D.-R. Yu, and Z.-X. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation," *J. Softw.*, vol. 19, no. 3, pp. 640–649, 2008.
[21] H. Jia *et al.*, "Spectral clustering with neighborhood attribute reduction based on information entropy," *J. Comput.*, vol. 9, no. 6, pp. 1316–1324, 2014.
[22] G. Wang *et al.*, "Decision table reduction based on conditional information entropy," *Chin. J. Comput.*, vol. 25, no. 7, pp. 759–766, 2002.
[23] S. Wu and P. Gou, "Attribute reduction algorithm on rough set and information entropy and its application," *Comput. Eng.*, vol. 37, no. 7, pp. 56–61, 2011.
[24] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.
[25] X. Zhang *et al.*, "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy," *Pattern Recognit.*, vol. 56, pp. 1–15, 2016.
[26] A. Calcagnì, L. Lombardi, and E. Pascali, "A dimension reduction technique for two-mode non-convex fuzzy data," *Soft Comput.*, vol. 20, no. 2, pp. 749–762, 2016.
[27] X. Zhang, J. Li, and H. Yu, "Local density adaptive similarity measurement for spectral clustering," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 352–358, 2011.
[28] H. Jia, S. Ding, and M. Du, "Self-tuning p-spectral clustering based on shared nearest neighbors," *Cognit. Comput.*, vol. 7, no. 5, pp. 622–632, 2015.

LIN LI was born in China in 1982. She received the B.S. degree from the Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China. She is currently an Associate Professor with the Department of Computer Science and Technology, China University of Mining and Technology. Her main research interests include clustering algorithm.

JIANHUA YUE was born in China in 1964. He received the Ph.D. degree from the College of Environmental Science and Engineering, Tongji University, Shanghai, China. He is currently a Professor and a Ph.D. Tutor with the College of Environmental Science and Engineering, China University of Mining and Technology, Xuzhou, China. His main research interests include resources and earth science.