Improved feature selection algorithm based on maximal nearest - neighbor rough approximation

Lin Lv
School of Information Science and Engineering,
Shandong Normal University
Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology
Jinan Shandong 250358, China
1069315857@qq.com

Yongqing Wei
Basic Education Department
Shandong Police College
Jinan Shandong 250014, China
18766175865@163.com

Min Ren
School of Mathematics and Quantity Economy,
Shandong University of Finance and Economics
School of Information Science and Engineering,
Shandong Normal University
Jinan Shandong 250014, China
rm_@163.com

Jing Yi
School of computer science and technology,
Shandong Jianzhu University
Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology
Jinan Shandong 250101, China
173802053@qq.com

***Abstract*—The feature selection algorithm based on maximal nearest neighbor rough approximation can not only deal with the mixed data, but also avoid the choice of the parameter values in the feature selection algorithm based on neighborhood rough sets. And it reduces the judgement of the sample. But the evaluation standard of this method only considers the importance of a single attribute which is relative to the result of the decision while calculating the importance of the attribute. It ignores the influence of the interaction between the attributes on the result of decision. So this paper sets up the new evaluation standard which is considered the influence of the attributes, and a forward greedy feature selection algorithm is constructed. Experiments show that the proposed algorithm can not only select fewer features, but also improve the accuracy of the classification.**

Keywords-feature selection; maximal nearest-neighbor; rough approximation; importance

I. INTRODUCTION

People often need to deal with data sets that contain a lot of features and a large number of examples in data analysis. In this class of data sets, some features are redundant or even irrelevant. The existence of redundant and irrelevant features will reduce the efficiency of the learning algorithm[1]. Therefore, it is necessary for us to preprocess the data to remove the redundant features and noise while analyzing the data sets. This will use feature selection.

The rough set theory[2] proposed by Z.Pawlak in Poland in 1982. It is a new mathematical tool for dealing with fuzzy and uncertain problems. At present, rough set has become a research hotspot in the field of artificial intelligence. It has been widely used in many fields such as machine learning, pattern recognition, intelligent control, decision analysis, knowledge acquisition, data mining and so on. Feature selection based on rough set method is to select a feature subset that is the same as the original feature set.

Searching and evaluation are two important feature selection steps. The most common search strategy is sequential search. And there are some heuristic methods, such as floating point search, column search, bidirectional search and gene search[3], and so on. The common search criteria contain distance metrics[4], dependency metrics[5-6], information metrics [7-8]. Also, there is a kind of search criteria based on rough sets[9]. However, the classical feature selection method is only suitable for dealing with discrete data. For continuous data, it needs to be discretized. But the process of discretization inevitably produces the information loss. So, a variety of feature selection methods suitable for continuous data are proposed. The feature selection algorithm in literature[10-11] is to construct a rough set model of neighborhood relation. For data of mixed type, Hu proposed a feature selection algorithm based on rough set. This method can deal with mixed data directly by setting the neighborhood parameter values. So it avoids the discretization of continuous data and improves the performance of the feature selection effectively. However, the algorithm needs to compute and preserve the neighborhood of the sample repeatedly. And the selection of the neighborhood of the sample is lack of the support of the theoretical model. The feature selection algorithm based on maximal nearest-neighbor rough approximation[12] calculates the neighborhood of the sample by a theoretical algorithm. It sets up a classified interval so that the neighborhood of the sample does not need to be calculated repeatedly. Compared with the original algorithm, the algorithm can select fewer features and improve the performance of classification. But it only considers the direct effect of a single attribute to the decision results in the calcu

978-1-5090-3906-7/16 $31.00 © 2016 IEEE
DOI 10.1109/ITME.2016.21

lation of the attribute results. The indirect effect of interaction between attributes is ignored. So this paper proposes an improved feature selection algorithm based on maximal nearest -neighbor rough approximation.

II. Rough Approximation Based On Maximal Nearest-neighbor

When we select the neighborhood of the sample in the neighborhood rough set model, it lacks the support of the theoretical model. For this problem, the feature selection algorithm based on maximal nearest-neighbor rough approximation is proposed.

Definition 1 Given space U of N dimension, Δ: $R^N * R^N \rightarrow R$,we call the Δ is a metric on R^N, if it meet:

(1)$\Delta(x_1, x_2) \geq 0$, $\Delta(x_1, x_2) = 0$, if and only if $x_1 = x_2$, $\forall x_1, x_2 \in R^N$;

(2)$\Delta(x_1, x_2) = \Delta(x_2, x_1)$, $\forall x_1, x_2 \in R^N$;

(3) $\Delta(x_1, x_3) \leq \Delta(x_1, x_2) + \Delta(x_2, x_3)$, $\forall x_1, x_2, x_3 \in R^N$;

So <U, Δ> is called metric space. Euclidean distance is commonly used in real space metric, it is also used in this paper.

Definition 2 Supposed <U, Δ> is non-metric space. U is a non-empty set. Δ is the distance function of U, x∈U, the maximum nearest-neighbor point set of X is

$$m(x) = \{y|\Delta(x, y) < d(x), y \in U\} \tag{1}$$

among the formula,

$$d(x) = \max\big(d_1(x),\ d_2(x)\big)$$
$$d_1(x) = \Delta(x - NH(x))$$
$$d_2(x) = \Delta(x - NM(x))$$

$NH(x)$ represents the similar sample that is closest to the X in the sample space. $NM(x)$ represents the different sample that is closest to the X in the sample space. $\Delta(x - NH(x))$ and$\Delta(x - NM(x))$ represent the distance of the sample point x to the $NH(x)$ and $NM(x)$.

Definition 3 Given a sample set $U = \{x_1, x_2, \cdots x_n\}$, C is a feature set that describes the D. U is a set of decision attribute. If a set of neighborhood relation is generated by C, the NDT=<U, C, D> is called a neighborhood decision system.

Definition 4 Given a neighborhood decision system NDT =<U, C, D>, D divides U into equivalent classes: $x_1, x_2, \cdots, x_n$, A⊆C generates neighborhood relation of R_A on the U, then the maximal nearest-neighbor lower approximation and upper approximation of the decision D is:

$$\underline{R_A}D = \left\{\underline{R_A}X_1,\ \underline{R_A}X_2,\ \cdots,\ \underline{R_A}X_N\right\} \tag{2}$$
$$\overline{R_A}D = \left\{\overline{R_A}X_1,\ \overline{R_A}X_2,\ \cdots,\ \overline{R_A}X_N\right\} \tag{3}$$
$$\underline{R_A}X = \left\{x_j|m_A\left(x_j\right) \subseteq Y, x_j \in U\right\} \tag{4}$$
$$\overline{R_A}X = \left\{x_j|m_A\left(x_j\right) \cap Y \neq \emptyset, x_j \in U\right\} \tag{5}$$

$m_A\left(x_j\right)$ is the maximal nearest neighbor information particle which is generated by attribute A and metric Δ. $\underline{R_A}D$ is also called positive domain. It is the set of objects that are fully contained with D in the maximal nearest neighbors. $\overline{R_A}D$ is the set of objects that are intersected with D but not empty in the maximal nearest neighbors.

Definition 5 classified interval: $margin(x) = d_2(x) - d_1(x)$

When margin(x)>0, $d_2(x)$>$d_1(x)$, the distance between the X and the different sample which is the closest to the X is greater than the distance between the X and the similar sample which is the closest to the X. The sample X can belong to the same decision class in its maximal nearest neighbor. And the sample X can be classified exactly. So it is easy to calculate the lower approximation. And it does not have to calculate whether the sample of the maximal nearest-neighbor belongs to the same decision class repeatedly.

Definition 6 Maximal nearest-neighbor dependency: Given MDT=<U, C, D>, the maximal nearest-neighbor dependency of the decision attribute D for condition attribute A is :

$$\gamma_A(D) = Card\left(\underline{R_A}D\right)/Card(U) \tag{6}$$

$0 \leq \gamma_A(D) \leq 1$, it represents the ratio of the samples that can be fully contained in a particular class of decision to the full samples according to the description of condition attribute A. From the formula, we can see that the value of the $\underline{R_A}D$ is greater, the dependence of the decision attribute D on the condition attribute A is stronger.

Definition 7 Given neighborhood decision system<U, C, D>, A ⊆ C, if the attribute A satisfies:

(1) $\forall a \in A$, $\gamma_{A-a}(D) < \gamma_A(D)$;

(2) $\gamma_A(D) = \gamma_C(D)$;

A is called the relative reduction in the maximal nearest-neighbor.

Definition 8 Attribute importance: Given MDT=<U, C, D>, a ∉A, then for the decision attribute D, the importance of the attribute a is:

$$Msig(a, A, D) = \gamma_{A\cup a}(D) - \gamma_A(D) \tag{7}$$

III. Improved Feature Selection Algorithm Based On Maximal Nearest-neighbor Rough Approximation

The above standard only considers the impact of the attribute on the category, and it does not consider the effect between attributes. So the evaluation standard is reset. It is:

$$Msig'(a_i, A, D) = Msig(a_i, A, D) + \sum_{a_j \epsilon A}\left[Msig\left(a_j, A \cup a_i, D\right) - Msig(a_j, A, D)\right] \tag{8}$$

The first item is to select the attribute that has the greatest dependence with target categories. It is also the most important attribute. The second item can be seen as an indirect effect. It is the importance of attribute a_i on the basis of the attribute set A. If $Msig\left(a_j, A \cup a_i, D\right) - Msig(a_j, A, D)$>0, it shows that after adding the attribute a_i in the attribute set A, the importance of other attributes in the selected attribute set is improved. It can be seen indirect importance of the attribute a_i.

The flow chart of the algorithm is shown in Fig 1.

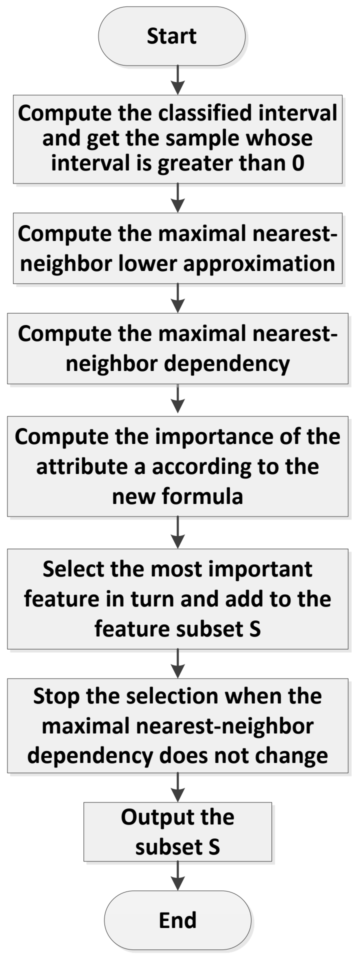

Figure 1

The specific steps of the algorithm are as follows:

Input: MDT=<U,A,D>

Output: a feature subset S

(1)Initialize the feature subset S.

(2)Compute the classified interval of sample and select the sa mple which the classified interval is more than 0. At this tim e, the samples of the X in the maximal nearest-neighbor belo ng to the same decision category. Those are the samples in a positive field.

(3)Compute the importance of attribute a according to the ne w formula. The maximal nearest-neighbor lower approximati on, upper approximation and dependency are computed by d efinition 4 and definition 6.

(4)Sort the attributes according to the importance. Select the feature that is the most important and add it to the subset S. I f Msig'(a,S,D)>0, it continues to carry on step(3). And then t he most important feature is selected and added to the subset S. The process is stopped when the value of the maximal nea rest-neighbor dependency no longer change.

(5)Output the feature selection subset S.

IV. EXPEIMENT AND RESULT ANALYSIS

A. Experimental Data

The test data is selected from UCI data set in the experi ment. Five sets of data are selected. The specific informatio n is shown in Table I.

TABLE I. EXPERIMENTAL DATA SET

dataset	number of example	number of feature	number of category
Wine	178	12	3
Zoo	101	17	7
Ionosphere	351	34	2
Sonar	208	60	2
Soybean	687	36	19

B. Experimental Results and Analysis

In order to verify the effectiveness of the proposed algor ithm, the algorithm in this paper is compared with the feature algorithm based on neighborhood rough set (NRS) and the f eature selection algorithm based on maximal nearest-neighbo r rough approximation(MNNRS). The algorithm is verified f rom the size of the selected feature subset and classified accu racy. The results are shown in Table II and Table III.

. The classified accuracy is defined as:

$$\text{Classified accuracy}=A\backslash B \tag{9}$$

A represents the number of the output classified results which is consistent with the actual classified result. B represe nts the total number of the sample.

TABLE II. THE FEATURE SELECTED RESULT OF FEATURE ELECTION ALGORITHM

dataset	NRS algorithm			MNNRS algorithm	algorithm in this paper
	0.10	0.12	0.14		
Wine	5	5	6	4	6
Zoo	9	8	9	8	7
Ionosphere	16	15	16	14	12
Sonar	6	6	7	6	6
Soybean	15	15	17	13	12

From the Table II, the three algorithms can effectively r educe the number of feature. The number of selected feature using the algorithm in this paper s is less than the other two a lgorithms. The number of selected feature using this algorith m in other two datasets is similar to the number of selected fe ature using other two algorithms. But the types of selected fe ature are different.

The neural network is used as classifier to test the algorit hms in the experiment. The results are shown in Table III.

TABLE III. THE ACCURACY OF CLASSIFICATION(%)

dataset	NRS algorithm			MNNRS algorithm	algorithm in this paper
	0.10	0.12	0.14		
Wine	97.2	94.4	97.2	95.6	97.1
Zoo	96.6	96.5	96.2	96.4	96.0
Ionosphere	93.9	94.1	94.5	94.4	94.8
Sonar	89.1	89.5	89.3	89.7	90.1
Soybean	94.1	94.3	93.9	94.5	94.6

From the Table III, the classified accuracy of the algorit hm in this paper is higher than that of other algorithms in mo st datasets. Although the accuracy is lower in some datasets, the gap is not big.

Overall, the algorithm in this paper has good performan ce in the selection of features and classified accuracy. It is ev en better than the other algorithms in some datasets.

V. Conclusion

According to the shortcomings of the feature selection al gorithm based on maximal nearest-neighbor rough approxim ation, the improved algorithm is proposed. It fully considers t he influence of the interaction between the attributes on the d ecision result. The evaluation standard is reset. And the forw ard search method is used for feature selection. Through testi ng on part of the dataset, we can see that the algorithm in this paper shows better performance compared with the feature s election based on the neighborhood rough approximation and the feature selection algorithm based on maximal nearest-nei ghbor rough approximation. The selected feature subset is m oderate in size. And the performance of the classification is g ood.

Acknowledgment

This work is partially supported by National Natural Scienc e Foundation of China (61373148, 61502151), Shandong Pr ovince Natural Science Foundation (ZR2012FM038, ZR201 4FL010), Shandong Province Outstanding Young Scientist Award Fund (BS2013DX033), Science Foundation of Minis try of Education of China(14YJC860042), Project of Shand ong Province Higher Educational Science and Technology P rogram (No.J13LN19, No.J15LN02), and Project of Shando ng Province university science and technology(NO.J15LN2 2).

References

[1] Xu Junling, Zhou yuming, Chen lin, et al, "An unsupervised feature selection approachbased on mutual information,"Journal of Computer Research and Development,vol.49, no.2, pp.372-382, 2012.)

[2] Pawlak Z. Rough Sets—Theoretical aspects of reasoning about Data. Dordrecht: Kluwer Academic, 1991.

[3] Dash M, Liu H, "Feature selection for classification," International Journal of Intelligent Data Analysis, vol.1, no.3, pp.131-156, 1997.

[4] Robnik-Sikonia M, Kononenko I, "Theoretical and empirical analysis of ReliefF and RReliefF," Machine Learning, vol.53. no.1, pp.23-69, 2003.

[5] Mitra P, Murthy C A, Pal S K, "Unsupervised feature selection using feature similarity," IEEE Trans on Pattern Analysis and Machine Intelligence, vol.24, no,3, pp. 301-312, 2002.

[6] Wei H L, Billings S A, "Feature subset selection and ranking for data dimensionality reduction," IEEE Trans on Pattern Analysis and Machine Intelligence, vol.29, no.1, pp.162-166, 2007.

[7] Battiti R, "Using mutual information for selecting features in supervised neural net learning," IEEE Trans on Neutral Networks, vol.5, no.4, pp. 537-550,1994.

[8] Peng H, Long F, Ding C, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy;" IEEE Trans on Pattern Analysis and Machine Intelligence, vol.27, no.8 pp.1226-1238, 2005.

[9] Guyon I, Elisseeff A, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol.3, no.1 pp.1157-1182, 2003.

[10] Qian Yuhua, Liang Jiye, Yao Yi yu, et al, "MGRS: A multi-granulation rough set," Information Sciences, vol.180, no.6 pp.949-970, 2010.

[11] Duan jie, Hu qinghua, Zhang lingjun, et al, "Feature selection for multi-label classification based on neighborhood rough sets," Journal of Computer Research and Development, vol.52, no.1, pp.56-65, 2015.

[12] Liu Jinghua, Lin Menglei, Wang Chenxi, et al, "Feature selection algorithm based on maximal nearest-neighbor rough approximation," Journal of Chinese Computer System, vol.36, no.8, pp.833-1836, 2015.