

Reduction of Variables through Nearest Neighbor Relations in Threshold Networks

Nohiro Ishii ^{*}, Kazunori Iwata [†], Kazuya Odagiri [‡]
Toyoshiro Nakashima [‡], Tokuro Matsuo ^{*}

Abstract

Threshold networks are useful as a fundamental technology in the recent learning and AI domains. Reduction of data variables in threshold networks is an important issue and it is needed for the processing of higher dimensional data in the application domains and AI. Boolean and rough set is fundamental and useful to reduce higher dimensional data to lower one for the classification. We develop a reduction of data variables and classification method based on geometrical reasoning, which is characterized by nearest neighbor relations. In this paper, the nearest neighbor relations are shown to be useful for the reduction of variables and their classification in threshold networks. The Boolean operation and convex cones generated by the nearest neighbor relations derive the reduced variables of data and the classifications using them. Then, the edges of convex cones are compared for the reduction of variables. Further, hyperplanes with reduced variables are generated on the same convex cones for data classification.

Keywords: Nearest neighbor relation, Boolean operation, Convex cones, Reduction of variables

1 Introduction

By Pawlak's rough set theory[1], a reduct is a minimal subset of features, which has the discernibility power as using the entire features, which shows the dimensionality reduction of features. Skowron[2,3] developed the reduct derivation by using the Boolean reasoning for the discernibility of data, which is a computationally complex task using all the data. A new consistent method for the generation of reduced variables and their classification in threshold networks is expected from the point of the efficient processing of data. In this paper, we have developed a method of reduction of data variables and the classification using the nearest neighbor relations[9], which are proposed to be relations with minimal distance between different classes. First, it is

^{*} Advanced Institute for Industrial Technology, Tokyo, Japan

[†] Aichi University, Nagoya, Japan

[‡] Sugiyama Jyogakuen University, Nagoya, Japan

shown that the nearest neighbor relations are useful to derive minimal Boolean function in threshold function[6,13,14]. Next, to derive the complete reducts, a geometric reasoning is developed, in which convex cones are generated based on the nearest neighbor relations in the linear subspaces[10,11,15]. Then, the degenerate convex cones and their operations using nearest neighbor relation, are developed. The dependent relations and algebraic operations of the edges on the degenerate convex cones in the linear subspaces are derived for the reduction of variables. Finally, the classification using the reduced variables is performed correctly based on the same convex cones of nearest neighbor relations.

2 Nearest Neighbor Relation in Threshold Function

The nearest neighbor relation is also applicable to the generation of threshold functions. The threshold function is a Boolean function based on the n -dimensional cube with 2^n vertices of n components of 1 or 0. The function f is characterized by the hyperplane $WX - \theta$ with the weight vector $W (= (w_1, w_2, \dots, w_n))$ and threshold θ . The X is a vertex of the cube 2^n . In the following, the threshold function is assumed to be positive and canonical threshold function, in which the Boolean variables hold the partial order[14].

Definition 2.1 The nearest neighbor relation (X_i, X_j) on the threshold function is defined to be vertices satisfying the following equation,

$$\{(X_i, X_j) : f(X_i) \neq f(X_j) \wedge |X_i - X_j| \leq \delta (= 1)\} \quad (1)$$

, where $\delta = 1$ shows one bit difference between X_i and X_j in the Hamming distance (also in the Euclidean distance).

The boundary vertex near the hyperplane, which realizes its threshold function as follows.

Definition 2.2 The boundary vertex X is defined to be the vertex which satisfies

$$|WX - \theta| \leq |WY - \theta| \quad \text{for the } X (\neq Y \in 2^n) \quad (2)$$

Theorem 2.3 The boundary vertex X becomes an element of nearest neighbor relation in the threshold function.

Theorem 2.4 The set of the element of the nearest neighbor relation is admissible set of threshold function.

Theorem 2.5 The vertices X_i and X_j in the nearest neighbor relation (X_i, X_j) are the adjacent vectors, each of which belongs to different class through the hyperplane.

This is proved, since the boundary vertex X is the nearest neighbor data to the hyperplane, which divides the true and the false data. This is shown in Figure 1. The nearest neighbor relation is not necessarily boundary vertex. Since the boundary vertices determine the hyperplane of the threshold function, the nearest neighbor relation also characterizes the threshold function. The data set is called to be admissible set of f , if the set realizes a threshold function f .

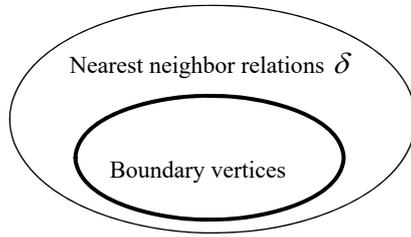


Figure 1: Boundary vertices and nearest neighbor relations in threshold function

Theorem 2.6 The nearest neighbor relations $\{(X_i, X_j)\}$ in a threshold function f is unique in f .

This theorem shows the line between adjacent vectors X_i and X_j with one bit difference is crossed by the hyperplane.

This is proved by the contradiction. Assume a threshold function has two nearest neighbor relations NNR_A and NNR_B , in which $(X_{iA}, X_{jA}) \neq (X_{iB}, X_{jB})$ holds. First, since X_{iB} is not as true vertex in the NNR_A , it exists as a false vertex. Then (X_{iB}, X_{jB}) does not make nearest neighbor relation of f , which contradicts the assertion. Second, when the partial ordering $X_{iA} \prec X_{iB}$ holds[14], X_{iB} does not make nearest neighbor relation of f with X_{jB} in one bit distance. Thus, $\{(X_i, X_j)\}$ is unique in the given threshold function f . Generation of threshold function is performed as follows. As an example, the three dimensional vertices are given as follows. As true valued vertices, (101), (110) and (111) are given. As false valued vertices, (000), (010), (100), (001), and (011) are given. Then, the 3-dimensional cube is shown in Fig.3, in which the black circle, \bullet belongs to +1 class, while the white circle, \circ belongs to 0 class. The black circle vertices are (101), (110) and (111), while the white circle vertices are (000), (001), (010), (100) and (011).

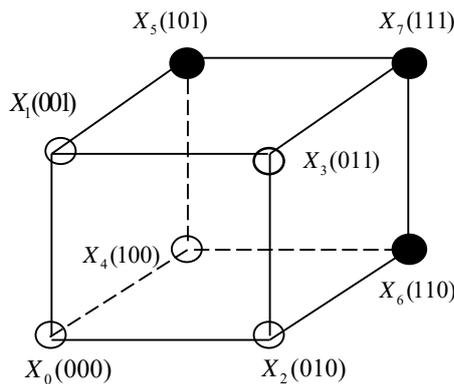


Figure 2: Example of a threshold function in 3-dimensional cube

2.1 Logical Operation for Nearest Neighbor Relation

Boolean reasoning of the nearest neighbor relations in threshold function are shown as follows. Directed arrow vector is introduced here for the nearest neighbor relations. In Figure 2, the true vertex (101) has nearest neighbor relations as $\{(101), (001)\}$ and $\{(101), (100)\}$, which are shown in Figure 3. Then, the directed arrow vector indicates the vector from the true vertex to the false one for the nearest neighbor relation, which is shown as $\overline{\{(101), (001)\}}$.

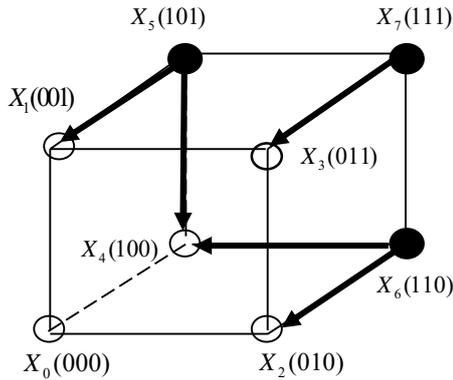


Figure 3: Directed arrows for nearest neighbor relations

Two directed arrow vectors, $\overline{\{(101), (001)\}}$ and $\overline{\{(101), (100)\}}$ generate one plane. This plane is represented in the Boolean AND operation, which is shown in Figure 4. The directed arrow vector has the x_3 variable between the true vertex and the false one, which is different between their vertices. Thus, $\overline{\{(101), (100)\}}$ has x_3 variable, while $\overline{\{(101), (001)\}}$ has x_1 variable. By the AND operation x_1 and x_3 , the Boolean product $x_1 \cdot x_3$ is generated. Similarly, different two directed arrow vectors, $\overline{\{(110), (100)\}}$ and $\overline{\{(110), (010)\}}$ generates another plane, which is shown in Figure 4. From this plane, the Boolean product $x_1 \cdot x_2$ is generated. Since these two planes are orthogonal, the Boolean OR operation is used to connect these two planes. The remained directed arrow vector in Figure 4, is $\overline{\{(111), (011)\}}$, which is included either planes as the vector, since the vector is parallel to each plane. Thus, by OR connecting two perpendicular planes, the threshold function $x_1 \cdot x_3 + x_1 \cdot x_2$ is obtained.

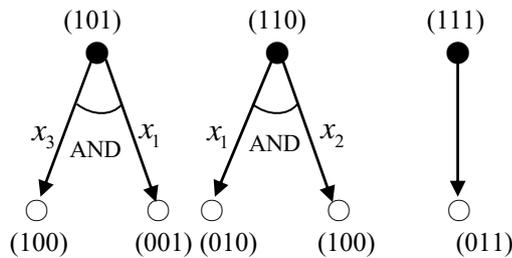


Figure 4: Boolean operations for nearest neighbor relations

2.2 Application of Logical Operation to Discernibility Matrix

The logical operations for nearest neighbor relations is applied to the discernibility matrix [3,4] to generate threshold function. The discernibility matrix of the nearest neighbor relation shows a necessary condition to be a threshold function. In Figure 2, a true valued vertex (101) has nearest neighbor relations as $\{(101), (001)\}$ and $\{(101), (100)\}$. The Boolean reasoning in these relations becomes a Boolean product in Table 4. The Boolean product, called minterm satisfies to be 1 for (101), while to be 0 for (001) and (100). Similarly, a true valued data (110) has nearest neighbor relations as $\{(110),(100)\}$ and $\{(110), (010)\}$. From these relations, the Boolean minterm is generated. Finally, the minterm with one variable is generated from the relation $\{(111),(011)\}$. The matrix shows different variables between adjacent vectors with different classes 1 or 0. The discernibility matrix of nearest neighbor data in Fig.3 is shown in Table 4. In Figure 2, a true valued data (101) has nearest neighbor relations as $\{(101), (001)\}$ and $\{(101), (100)\}$ as shown in shaded cells in Table 4. In Table 4, the difference of variables between $\bullet(101)$ and $\circ(001)$ is x_1 . Similarly, the difference between (101) and (100) is x_3 . For the Boolean realization of the difference of these variables is performed by the Boolean product $x_1 \cdot x_3$.

Table 4 : Discernibility matrix of Figure 2 data

	● (101)	● (110)	● (111)
○ (001)	x_1
○ (100)	x_3	x_2	...
○ (010)	...	x_3	...
○ (011)	x_1

The Boolean product, called minterm $x_1 \cdot x_3$ satisfies to be 1 for (101), while to be 0 for (001) and (100). Similarly, a true valued data (110) has nearest neighbor relations as $\{(110),(100)\}$ and $\{(110), (010)\}$. From these relations, the Boolean minterm is generated. Finally, the minterm with one variable is generated from the relation $\{(111),(011)\}$. But, this minterm is removed from the proposed indiscernibility matrix in Table 5, in which the sufficiency condition of these minterms are checked for a threshold function by taking common variables

Table 5 : Indiscernibility matrix of Figure 2 data

	● (101)	● (110)	● (111)
○ (001)	$x_2 \cdot x_3$...	x_3
○ (100)	$x_1 \cdot x_2$	$x_1 \cdot x_3$	x_1
○ (010)	...	$x_2 \cdot x_3$	x_2
○ (011)	x_3	x_3	$x_2 \cdot x_3$

between true and false vertices. The common variables x_1 and x_2 are taken as a Boolean product in the cell between $\bullet(101)$ and $\circ(001)$, which is not used for the difference (101) and (001) . But, the minterm derived in Table 4, $x_1 \cdot x_2$, discriminates the true vector (101) and the false vector (001) . Finally, the minterm $x_1 \cdot x_3$ is generated from the relation $\{(111), (011)\}$ in Table 4 in the shaded cell. The common variable x_1 also exists in Table 5, between true vector (111) and false vertex (100) . But, to realize the difference between (111) and (100) , the minterms $x_1 \cdot x_3$ or $x_1 \cdot x_2$ is applicable to this cell. Thus, the minterm $x_1 \cdot x_2$ derived in Table 4, which is only one common variable in Table 5 is removed by checking the discernibility and the indiscernibility matrices. Then, the minterm x_1 is removed from the Boolean sum from Table 4. Thus, the Boolean function obtained is

$$f = x_1 \cdot x_2 + x_1 \cdot x_3 \quad (3)$$

The Boolean function f becomes a threshold function, since a hyperplane exists to satisfy the equation (3). The Nearest neighbor relations in threshold function are minimal information for the discrimination by the hyperplane. Generally, this is described in the following theorem.

Theorem 2.7 Nearest neighbor relations in the threshold function are minimal information for generating the given threshold function in the Boolean logical form.

Nearest neighbor relations are extended to the reduction of variables with general values for the classification in the threshold networks in the next section.

3 Extension to Threshold Networks Based on Nearest Neighbor Relations

As the application stage from the threshold function, a nearest neighbor relation with minimal distance is introduced here for the extension to threshold networks. The relation with minimal distance plays an important role for the reduction of variables for threshold networks.

Definition 3.1 A nearest neighbor relation with minimal distance is a set of pair of instances, which are described in

$$\{(x_i, x_j) : d(x_i) \neq d(x_j) \wedge |x_i - x_j| \leq \delta\} \quad (4)$$

, where $|x_i - x_j|$ shows the distance between x_i and x_j . Further, $d(x_i)$ is a decision function and δ is the minimal distance. Then, x_i and x_j in the equation (1) are called to be in the x_j nearest neighbor relation with minimal distance δ .

Lemma 3.2 Respective Boolean term consisting of the set $\{nn_{ij}\}$ becomes a necessary condition to be reduced in the Boolean expression.

This is trivial, since the product of respective Boolean term becomes reduced in the Boolean expression.

Lemma 3.3 Boolean product of respective terms corresponding to the set $\{nn_{ij}\}$ becomes a necessary condition to be reducts in the Boolean expression.

Lemma 3.4 Reducts in the Boolean expression are included in the Boolean term of Lemma 3.2 and the Boolean product in Lemma 3.3.

Figure 2 shows that nearest neighbor relation with classification is a necessary condition in the Boolean expression for reducts, but not sufficient condition. The distance δ of the nearest neighbor relation in the equation (4) is compared with the distance δ' of the relation in the following theorem.

Theorem 3.5 If the distance δ is greater than the δ' , i.e., $\delta > \delta'$ in the equation (3), the Boolean expression of the case of δ' includes that of δ .

This is by the reason that the Boolean expression of the nearest neighbor relation is consists of the Boolean product of variables of the relation. The number of variables in the distance δ' are less than that of δ . Thus, the nearest neighbor relation with distance δ' includes the ellipse of δ in Figure 5.

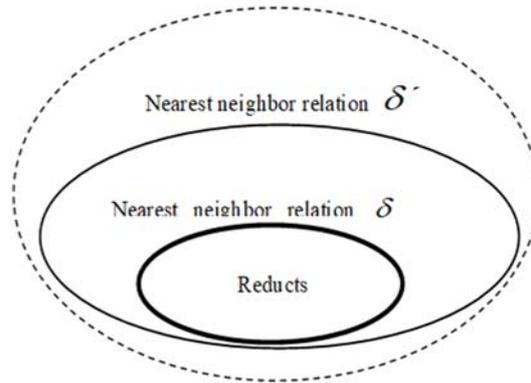


Figure 5: Boolean condition of nearest neighbor relations and reducts

To find minimal nearest neighbor relations, the divide and conquer algorithm is applied, which is well known to search the closest pair[7,12]. A schematic figure of the search for closest pair of the set(data) is shown in Figure 6. The closest pair of points is derived as follows. Divide the set(data instances) into equal sized parts by the line l and recursively compute the minimal distance in each part, where d is the minimal of the two minimal distances[7]. Sort the remaining points according to their y coordinates. Merge the two sorted lists into one sorted list. These dividing steps takes $O(\log n)$. Eliminate the points that lie farther than d apart from l and n is the total number of data. Thus, the divide and conquer algorithm derives $O(n \log n)$ for the operations to search the nearest neighbor relations. The closest pair of points is derived as follows. Divide the set(data instances) into equal sized parts by the line l and recursively compute the minimal distance in each part, where d is the minimal of the two minimal distances[7]. Sort the remaining points according to their y coordinates. Merge the two sorted lists into one sorted list. The divide and conquer algorithm[7,12] derives $O(n \log n)$ for the operations to search the nearest neighbor relations.

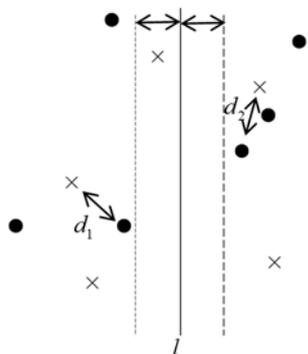


Figure 6: Search of nearest neighbor relations among data

Theorem 3.6 When the number of the instances is n , the number of the nearest neighbor relations is estimated to be $O(n)$ and the operations finding the nearest neighbor relations are estimated to be $O(n \log n)$ in the complexity order.

4 Reduction of Variables in Threshold Networks

As the dimensionality reduction, we adopt the reduct, which is developed from rough set theory. Decision table is necessary for the representation between data and its attributes. An example of the decision table is shown in Table 6. The left side data in the column in Table 1 as shown in, $\{x_1, x_2, x_3, \dots, x_7\}$ is a set of instances, while the data $\{a, b, c, d\}$ on the upper row, shows the set of attributes of the instance. The contents of each row in Table 6 shows numeral values of the corresponding instance. In case of instance x_1 in Table 6, the value of the attribute a, is $a(x_1)=1$. That of the attribute b, is $b(x_1)=0$. Since $a(x_1)=1$ and $a(x_5)=2$, $a(x_1) \neq a(x_5)$ holds. In Table 6, Boolean variables of the nearest neighbor relations are shown in the gray elements, which are derived in Table 6. Linearly separated spaces play an important role for the dimensionality reduction for reducts.

Table 6: Decision table of data example(instances)

Attribute	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	class
x_1	1	0	2	1	+ 1
x_2	1	0	2	0	+1
x_3	2	2	0	0	-1
x_4	1	2	2	1	-1
x_5	2	1	0	1	-1
x_6	2	1	1	0	+1
x_7	2	1	2	1	-1

First, the instance space is divided by piecewise linear classification subspaces. Their divided subspaces generate convex cones by their nearest neighbor relations. In Figure 7, linearly separated classifications are shown for the data in Table 6. As the first step, One hyperplane A_1 divides the instances of data $\{x_3, x_5, x_7, x_4\}$ shown with \times from those of $\{x_2, x_6\}$

shown with \bullet . But, it misclassifies the instance $\{x_1\}$ with \bullet from $\{x_3, x_5, x_7, x_4, x_1\}$ with \times . As the second step, another hyperplane A_2 divides the instance $\{x_1\}$ with \bullet from $\{x_3, x_5, x_7, x_4\}$ with \times . In the linearly separated subspaces, the divided instances are represented by the system of the linear inequality equations. Then, from Ky Fan's inequality theory[5], the inequality system composed of some independent equations and all the other inequality equations[5] as follows. Let the divided instances be p in the number and they are represented by the p inequalities in the following equation,

$$f_i(x) \geq \alpha_i \quad (1 \leq i \leq p) \quad (5)$$

,where $f(x) = 0$ is the hyperplane separating the divided instances. By Key Fan's theory[8], there exist the following r independent equations, whose solution satisfies the all the inequalities (5).

$$f_{\nu k}(x) = \alpha_{\nu k} \quad (1 \leq k \leq r) \quad (6)$$

,where $f_{\nu 1}, f_{\nu 2}, \dots, f_{\nu r}$ are independent equations, which are generated from the r independent instances. The hyperplane A_1 in Figure 7 is computed from independent instances of the nearest neighbor relations. Solution of the equation (6) becomes to be

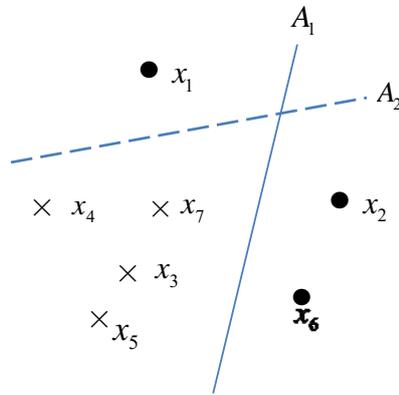


Figure 7: Piecewise linear separated classification for data in Table 6

$w_1 = (3/2)\varepsilon, w_2 = -2\varepsilon, w_3 = 0, w_4 = -2\varepsilon$. Thus the hyperplane A_1 separating the $\{x_3, x_5, x_7, x_4\}$ and $\{x_2, x_6\}$ becomes

$$-a \cdot \varepsilon - b \cdot \varepsilon + (3/2) \cdot c \cdot \varepsilon - d \cdot \varepsilon = 0 \quad (7)$$

The hyperplane A_1 separates instances $\{x_3, x_5, x_7, x_4\}$ and $\{x_2, x_6\}$. Among these instances, nearest neighbor relations $\{(x_3, x_6), (x_5, x_6), (x_7, x_6)\}$ hold. The hyperplane A_2

separates the instances $\{x_3, x_5, x_7, x_4\}$ and $\{x_1\}$ as shown in Figure 7. Among these instances, a nearest neighbor relation $\{(x_1, x_7)\}$ holds. From the hyperplane, A_1 the nearest neighbor relations $\{x_3, x_6\}$, $\{x_5, x_6\}$ and $\{x_6, x_7\}$ are made. The $\{x_3, x_6\}$ shows different values in their $\{b, c\}$ components in Table 1. Similarly, $\{x_5, x_6\}$ and $\{x_6, x_7\}$ show those values in $\{c, d\}$ components. To make complete reducts, the processing steps by geometrical analysis based on the nearest neighbor relations are shown in Figure 8.

Processing steps for the reduction of variables are classified to the following three steps, which are shown in circles ① and ② in Figure 8.

Step ① Variables of nearest neighbor relations

Nearest neighbor relations generate convex cones, which are fundamental and reduced variables for the classification of instances. Each approximated reduct based on the convex cone is generated based on the nearest neighbor relation, in the step, ① in Figure 8. The obtained convex cone plays for the reduction of variables in steps ②

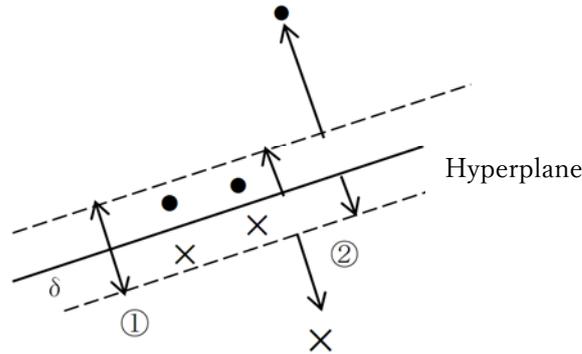


Figure 8: Processing steps based on hyperplane derived from nearest neighbor relation

Step ② Removal of variables using nearest neighbor relations

The convex cones based on the nearest neighbor relations are transformed to the reduced ones for the reduction of variables. Dependent relations and edge processing between the convex cones are applied for the complete reducts, which show the reduction of variables as the step ② in Figure 8. In the next section, the reduction of variables on the convex cones is shown in the next section 5, step (2.1) and step (2.2).

5 Geometrical Analysis Using Convex Cones

The geometrical reasoning approach for the generation of reducts is developed by generating convex cones [7,8] in the subspaces.

5.1 Generation of Convex Cone Based on Nearest Neighbor Relations

As the linear classification, two hyperplanes are computed in Figure 7 for data in Table 6. The hyperplane A_1 separates the data $\{x_3, x_5, x_7, x_4\}$ and $\{x_2, x_6\}$. From these data, convex cone is generated, firstly. We use capital letter notations $\{X_i\}$ for convex cones instead of small letters $\{x_i\}$ for instances in Table 6. The triangular convex cone is made from the nearest neighbor relation data in Table 6 as shown in Figure 8. The reference data X_6 is assigned, which is the data in the class +1.

The data $\{X_3, X_5, X_7\}$ with nearest neighbor relation in Table 6 generates a triangular convex cone in Figure 9. This is interpreted as follows. The data $\{X_3, X_5, X_7\}$ exists in the class -1, while the data $\{X_6\}$ is in the class +1 as the reference point. The vectors \vec{X}_3, \vec{X}_5 and \vec{X}_7 are generated as $\vec{X}_3 = X_3 - X_6$, $\vec{X}_5 = X_5 - X_6$ and $\vec{X}_7 = X_7 - X_6$ in the affine subspace. Any inner vector \vec{Y} in the triangular convex cone is indicated as

$$\vec{Y} = \alpha \vec{X}_3 + \beta \vec{X}_5 + \gamma \vec{X}_7 \quad (8)$$

,where $\alpha, \beta, \gamma \geq 0$ hold by convex condition. In the vectors in equation (8), By removing the reference point data, from vectors in equation (8), the following equation holds in the linear subspace,

$$Y = \alpha X_3 + \beta X_5 + \gamma X_7 \quad (9)$$

By applying the equation (8) removing reference point data, X_4 does not satisfy the equation (9) in the linear subspace, that is,

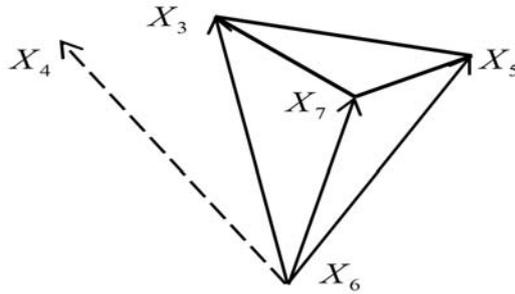


Figure 9: Convex cone by data $\{X_3, X_5, X_7\}$ with nearest neighbor relation generated on the hyperplane

$$X_4 \neq \alpha X_3 + \beta X_5 + \gamma X_7 \quad (10)$$

Then, the data X_4 constructs other convex cone as in Figure 9. Next, the reference data in the class +1 is changed to the reference data X_1 from that of X_6 in Figure 9. The Boolean variables derived from the nearest neighbor relations $(X_6, X_3), (X_6, X_5)$ and (X_6, X_7) on the triangular convex cone in Figure 9 are given as Boolean variables $\{b, c\}, \{c, d\}$ and $\{a, b\}$, respectively. Thus, the approximated reducts, $\{bc, bd, ac\}$ are obtained.

5.2 Data Chaining Degenerate of Nearest Neighbor Relations

To derive complete reducts, a degenerate convex cone is defined by the nearest neighbor relations. As shown in Figure 9, the convex is made of the reference data X_6 and data X_3, X_5, X_7 . From these edge data on the convex cone, the union of Boolean variables $\{b, c\}$ and $\{c, d\}$ is made, which consists of three variables (b, c, d) . Then, the degenerate convex cone is made from three variables (b, c, d) , which is removed the variable (a) as shown in Figure 10. Then, the degenerate convex cone in Figure 10 consists of X'_3, X'_5, X'_7 and X'_6 , which correspond to Boolean variables (b, c, d) .

5.3 Dependent Relations on Degenerate Convex Cones, Step (2.1)

Independent vectors of nearest neighbor relations, which consist of the degenerate convex in the affine subspaces, are useful for the operations of dimensional reduction. Independent vectors are derived in the affine subspaces[5,6,7] as follows.

$$\vec{X}'_3 = X'_3 - X'_6, \quad \vec{X}'_5 = X'_5 - X'_6 \quad \text{and} \quad \vec{X}'_7 = X'_7 - X'_6 \quad (11)$$

The vector X'_i is composed of components of attributes (b, c, d) of nearest neighbor relations. The dependent relation of these independent vectors removes the corresponding item on the indiscernibility matrix.

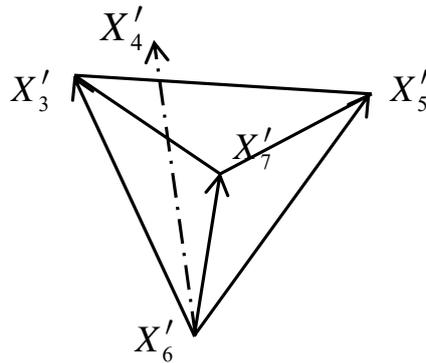


Fig. 5. Degenerate convex cone generated from X'_3, X'_5, X'_7 and X'_6

Since $X'_6 = (110)$, $X'_3 = (200)$, $X'_5 = (101)$, $X'_7 = (121)$ and $X'_4 = (221)$, the following linear equation is checked whether the edge value $(X_4 - X_6)'$ is dependent or not to independent vectors \vec{X}'_3, \vec{X}'_5 and \vec{X}'_7 .

$$\begin{pmatrix} X'_4 - X'_6 \\ 1 \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} \vec{X}'_3 \\ -1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} \vec{X}'_5 \\ -1 \\ 1 \end{pmatrix} + \gamma \begin{pmatrix} \vec{X}'_7 \\ 0 \\ 1 \end{pmatrix} \quad (12)$$

Further, we assume the vector $(X''_4 - X''_6)$, which composes of components (b, c) .

$$\begin{pmatrix} X''_4 - X''_6 \\ 1 \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} \vec{X}''_3 \\ 1 \\ -1 \end{pmatrix} + \gamma \begin{pmatrix} \vec{X}''_7 \\ 0 \\ 1 \end{pmatrix} \quad (13)$$

,where $\alpha=1$ and $\gamma=2$ hold. Thus, $(X''_4 - X''_6)$ is dependent to \vec{X}''_3 and \vec{X}''_7 . Similarly, $(X''_4 - X''_6)$ with components (c, d) is dependent to \vec{X}''_7 . Then, the corresponding Boolean term to $(X''_4 - X''_6)$ is removed. Generally, the following theorem holds.

Theorem 5.1 If a linear dependent equation of the vector value holds on the degenerate convex of the nearest neighbor relations, the corresponding Boolean term to the vector value is removed by the absorption of Boolean terms of the nearest neighbor relations.

5.4 Chaining of Edges between Degenerate Convex Cones, Step (2.2)

We can construct two degenerate convex cones as shown in Figure 11. The left side degenerate convex cone is made of the nearest neighbor relations which is from different classes data, while the right side one is made of the same class data.

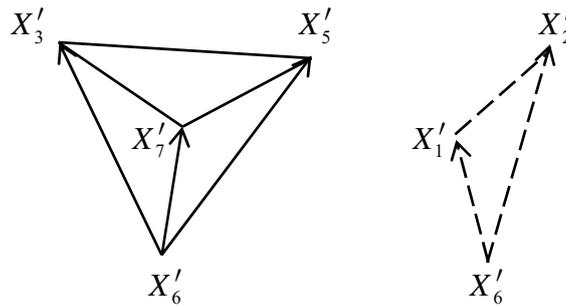


Figure 11: Degenerate convex cones: the left cone is from the nearest neighbor relations and the right one is from the same class data

The data difference equations except the nearest neighbor relations are shown in (14)

$$(X'_i - X'_j)_k \neq 0 \text{ for } i = 3, 4, 5, j = 1, 2 \text{ and } k = b, c, d \quad (14)$$

If the equations (14) hold for both $k = b$ and $k = c$, the element (x_i, x_j) in the discernibility matrix [11], is removed by the equality equation of $(X_i = X_j)_{k=b, k=c}$, since it is absorbed from the nearest neighbor relation $\{b, c\}$. Similarly, if the equations (14) holds for both $k = b$ and $k = c$, the element (x_i, x_j) is removed, since it is absorbed from the nearest neighbor rela

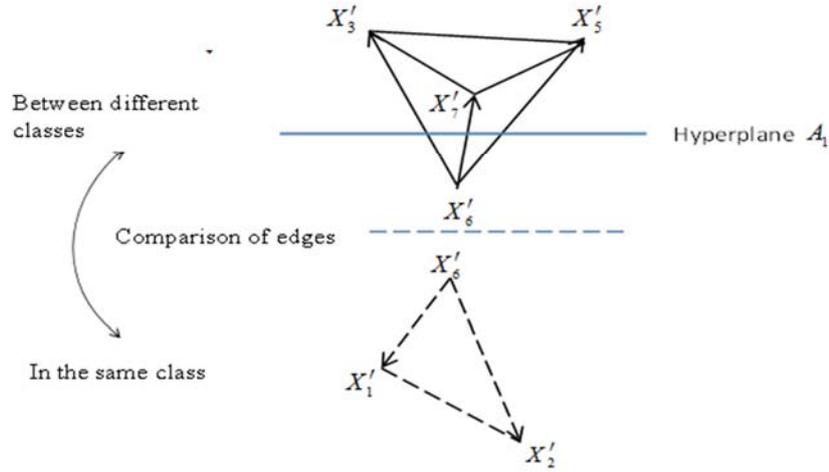


Figure 12: Comparison of edges on degenerate convex cones in the affine subspaces

tion $\{c, d\}$. This is interpreted in the following data chaining of the nearest neighbor relations. The left side of the equation (14) is replaced to the following equation using the nearest neighbor relation in Figure 12.

$$(X'_i - X'_j)_k = (X'_i - X'_6)_k - (X'_j - X'_6)_k \quad (14)$$

From the equation (12), the equation (11) becomes

$$(X'_i - X'_6)_k \neq (X'_j - X'_6)_k \quad (15)$$

The upper cone is constructed from the different classes data, while the lower cone is constructed from the same class data. Thus, the algebraic operation in the equation (15) compares the values between the upper convex cone and the lower one in Figure 12.

The left side of the equation (15) is the difference data between data in the different classes of the nearest neighbor relations, while the right side of the equation (15) is the difference data between data in the same class. These algebraic operations are generally described as the following theorem.

Theorem 5.2 The element (x_i, x_j) with Boolean variables are removed by the nearest relations with reference data X'_ξ , where X'_ξ is in the class $\{j\}(\neq \{i\})$, if the following equations hold in Figure 11

$$(X'_i - X'_\xi)_k \neq (X'_j - X'_\xi)_k \quad (16)$$

for $i \in \{i\}, j \in \{j\}$, i and j are in the different class and k is components in X'_i .

By Theorem 5.2, an example of $X'_\xi = X'_6$ and $(X'_3 - X'_2)$, which are in the case of $i = 3, j = 2$, is developed as follows,

$$(X'_3 - X'_2) = (X'_3 - X'_6) - (X'_2 - X'_6) \quad (17)$$

Since $(X'_3 - X'_6)_b = +1$ and $(X'_2 - X'_6)_b = -1$ hold, $(X'_3 - X'_6)_b \neq (X'_2 - X'_6)_b$. Similarly, $(X'_3 - X'_6)_c \neq (X'_2 - X'_6)_c$ holds. Thus, the element (x_3, x_2) with Boolean variables (b, c) is removed. Thus, the Boolean variables of the element $(x_i, x_j) \equiv (x_3, x_2)$ is removed by the absorption of the nearest neighbor relation. Similarly, $(x_i, x_j) \equiv (x_5, x_1), (x_3, x_2), (x_5, x_2)$ and (x_7, x_2) are removed. Similar operations $i = 3, 4, 5$ and $j = 1, 2$ are carried out.

Corollary 5.3 The element (x_i, x_j) with Boolean variables is not removed, if the equation (16) does not hold. Then, the Boolean sum of the variables in the (x_i, x_j) is multiplied to other Boolean sum of variables of the nearest neighbor relations.

Another approach can be adopted for the analysis in case of $\{x_1\}$. The hyperplane A_2 in Figure 7 was computed for the separation of data $\{x_1\}$ from $\{x_3, x_5, x_7, x_4\}$. As in the case of $j = 1$, X'_1 , a convex cone is generated on the relation (X_1, X_7) in Figure 13.

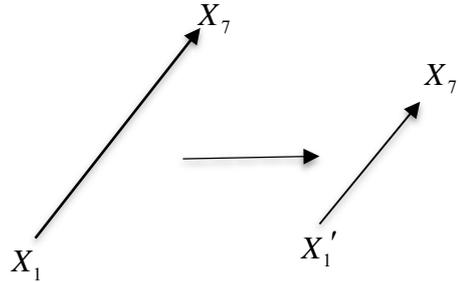


Figure 13: Convex cone generated based on hyperplane A_2

The Boolean variable (a, b) is generated on the convex cone (X_1, X_7) in Fig.8 based on the hyperplane A_2 . Degenerate convex cone (X'_1, X'_7) is generated in Fig.8, in which the edge with components a and b . In Figure 12, based on the degenerate convex cone (X'_1, X'_7) , a dependent relation exist between (X'_1, X'_3) and (X'_1, X'_7) . Similarly, a dependent relation

holds between (X'_1, X'_5) and (X'_1, X'_7) . But, the dependent relation does not hold between (X'_1, X'_4) and (X'_1, X'_7) , in which only Boolean variable (b) is derived in Figure 14. The Boolean variable (a, b) on the degenerate convex cone in Figure 12 removes (X'_3, X'_1) and (X'_5, X'_1) . However, the variable (b) in (X'_4, X'_1) is remained.

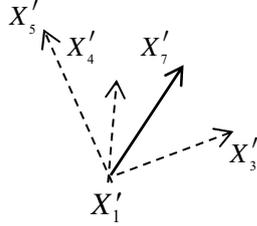


Figure 14: Dependent relations on degenerate convex cones

Thus, the variable (b) is selected by the hyperplane classification A_2 . Boolean sum terms $(b + c)$ and $(c + d)$ are derived based on the hyperplane A_1 and a term (b) on the hyperplane A_2 . By product terms, complete reducts become $\{bc, bd\}$.

Theorem 5.4 Dimensionality reduction of variables for reducts is realized by convex cones on the nearest neighbor relations, which are generated by the linear subspaces.

6 Classification of Reduced Variables in Threshold Network

The classification of instances is realized using linear classifiers with reduced variables, which are derived based on the geometrical reasoning in the previous section 5. The reduced variables for reducts are obtained based on the convex cones of the ne

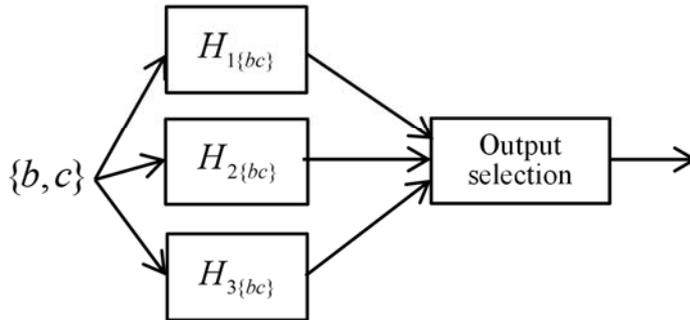


Figure 15: Hyperplanes generated on convex cones with nearest neighbor relations

est neighbor relations in Figure 9. Under the restricted conditions of inequalities with four variables of instances in Table 6, the relation of the hyperplane with the reduced variables are expected to be made clear. Introducing slack variables in the reduced ones of the inequalities, we can derive the hyperplanes with reduced variables. In the section 5, the reduction of variables are carried out, based on the geometric reasoning on convex cones of nearest neighbor relations. As the reduced variables, a reduct $\{bc\}$ of variables b and c is obtained

in the section 5.4. By using this reduct $\{bc\}$, a classification of the reduced variables is performed as shown in Figure 15. In Figure 15, three hyperplanes, $H_{1\{bc\}}$, $H_{2\{bc\}}$, $H_{3\{bc\}}$ are shown. The hyperplane A_1 in Figure 7 is shown in the equation (4), which is also in the degenerated convex in Figure 10. To compute the hyperplane $H_{1\{bc\}}$ with reduced two variables b and c , the hyperplane A_1 with four variables, a, b, c and d in the equation of the hyper plane A_1 is started,

$$A_1: \quad (3/2)a - 2b - 2d = 0 \quad (18)$$

, where $\varepsilon=1$ in the equation (7) is set. Since the reduced variables $\{b, c\}$ consists of two variables, three vertices x_3, x_6, x_7 on the convex in Figure 9 are taken on the convex in Figure 9 as the first reduced hyperplane, $H_{1\{bc\}}$. For the instance, x_3 , the following equation(19) holds using the hyperplane A_1 . Since $x_3 = (2, 2, 0, 0)$ belongs to -1 class,

$$(3/2) \cdot 2 - 2(1 + \Delta\xi_1) \cdot 2 + \Delta\xi_2 + \theta \leq 0 \quad (19)$$

, where $\Delta\xi_1$ and $\Delta\xi_2$ are slack variables. From the equation (19), we set

$$-4\Delta\xi_1 + \theta = -1 \quad (20)$$

Similarly, since $x_6 = (2, 1, 1, 0)$ belongs to +1 class,

$$(3/2) \cdot 2 - 2(1 + \Delta\xi_1) \cdot 1 + \Delta\xi_2 \cdot 1 - 2 \cdot 0 + \theta > 0 \quad (21)$$

Then, we set

$$-2\Delta\xi_1 + \Delta\xi_2 + \theta = +1 \quad (22)$$

Similarly, since $x_7 = (2, 1, 2, 1)$ belongs -1 class,

$$(3/2) \cdot 2 - 2(1 + \Delta\xi_1) \cdot 1 + \Delta\xi_2 \cdot 2 - 2 \cdot 1 + \theta \leq 0 \quad (23)$$

Then, we set

$$-2\Delta\xi_1 + 2\Delta\xi_2 + \theta = 0 \quad (24)$$

From equations (20), (22) and (24), $\Delta\xi_1 = (+3/2)$, $\Delta\xi_2 = -1$ and $\theta = +5$ are obtained.

Thus, from equations (19),(21) and (23), a hyperplane with reduced variables $\{b, c\}$ is given by coefficient products of each slack variables

$$-2 \cdot (+3/2)b - 1 \cdot c + 5 = -3b - c + 5 = 0 \quad (25)$$

The equation (25) of the hyperplane with reduced variables $\{b, c\}$, classifies to the respective classes of x_1, x_2, x_3, x_4, x_6 and x_7 , but it does not satisfy the class of x_5 . Then, we need another hyperplanes different from $H_{1\{bc\}}$. From another three vertices x_5, x_3, x_6 on the convex in Figure 9 as the second reduced hyperplane, $H_{2\{bc\}}$ are computed, which classifies to the respective classes of x_1, x_2, x_3, x_5 and x_6 , but it does not satisfy the classes of x_4 and x_7 . Similarly, from another three vertices x_5, x_3, x_7 on the convex in Figure 9 as the third reduced hyperplane, $H_{3\{bc\}}$ are computed, which classifies to the respective classes of x_3, x_4, x_5 and x_7 , but it does not satisfy the classes of x_1, x_2 and x_6 . All the correct classifications of the classes are carried out by majority voting using these three hyperplanes with reduced variables, $H_{1\{bc\}}, H_{2\{bc\}}, H_{3\{bc\}}$ in Figure 15. Then, all the instances are classified, correctly. Combining these three hyperplanes with reduced variables, another correct classification is also performed using the output selection by the comparison of their values.

7 Conclusion

In this paper, the reduction of data variables and the classification through the nearest neighbor relations are proposed in the threshold networks. For the threshold function, the method of reduction of variables is characterized by the Boolean operations of vectors of the nearest neighbor relations. For the data with general values in threshold networks, reduction of variables are realized based on convex cones made of the nearest neighbor relations in threshold networks. It is shown that the nearest neighbor relations derives approximated reducts of reduced variables. To derive complete reducts based on the approximated reducts, the degenerate convex cones are generated in the linear subspaces based on the nearest neighbor relations. Then, the dependent relations and the algebraic operations of edges on the degenerate convex cones are developed in the linear subspaces. The classification using the reduced variables is realized in the threshold networks.

References

- [1] Z. Pawlak, "Rough Sets," International Journal of Computer and Information Science, vol.11, 1982, pp.341-356.
- [2] Z. Pawlak and R. Slowinski, "Rough Set Approach to Multi-attribute Decision Analysis," European Journal of Operations Research 72, 1994, pp.443-459 .

- [3] A. Skowron and C. Rauszer, “The Discernibility Matrices and Functions in Information Systems,” in *Intelligent Decision Support- Handbook of Application and Advances of Rough Sets Theory*, pp.331-362, Kluwer Academic Publishers, Dordrecht, 1992
- [4] A. Skowron and L. Polkowski, “Decision Algorithms, A Survey of Rough Set Theoretic Methods,” *Fundamenta Informatica*, 30/3-4, pp. 1997, 345-358.
- [5] Ky Fan: *On Systems of Linear Inequalities, Linear Inequalities and Related Systems*, edited by H. W. Kuhn and A.W. Tucker, Princeton University Press, 99-156(1966)
- [6] T. M. Cover and P.E. Hart, “Nearest Neighbor Pattern Classification,” *IEEE Transactions on Information Theory*, Vol.13, No.1, 1967, pp.21-27.
- [7] F.P. Preparata and M.I. Shamos, *Computational Geometry*, Springer Verlag, 1993
- [8] W. Prenowitz and J.Jantosciak, *Join Geometries, A Theory of Convex Sets and Linear Geometry*, Springer Verlag, 2013,
- [9] N. Ishii, I. Torii, K. Iwata, K.Odagiri, T. Nakashima: *Generation and Nonlinear Mapping of Reducts-Nearest Neighbor Classification*. Chapter 5 in *Advances in Combining Intelligent Methods*, Springer Verlag, 93-108(2017)
- [10] N. Ishii, I.Torii, K. Iwata, K. Odagiri, T. Nakashima: *Generation of Reducts Based on Nearest Neighbor Relations and Boolean Reasoning*, HAIS2017, LNCS vol.10334, Springer, 391-401(2017)
- [11] N. Ishii, I. Torii, N. Mukai, K. Iwata and T. Nakashima, “Generation of Reducts and Threshold Function Using Discernibility and Indiscernibility Matrices”, *Proc. ACIS-SERA IEEE Comp. Soc.*, 55-61(2017)
- [12] A.V.Levitin, *Introduction to the Design and Analysis of Algorithms*, Addison Wesley, 2002
- [13] A. De, I. Diakonikolas, V. Feldman, R.A. Servedio, “Nearly Optimal Solutionms for the Chow Parameters Problem and Low-weight Approximation of Halfspaces”, *J.ACM*, Vol.61,No.2, 2014, pp.11:1-11:36.
- [14] S.T.Hu, *Threshold Logic*, University of California Press,1965
- [15] N. Ishii, I. Torii, N. Mukai, K. Iwata and T. Nakashima, “Incremental Reducts Based on Nearest Neighbor Relations and Linear Classification”, *Proc. IIAI-SCAI IEEE Comp. Soc.*, 528-533(2019)