TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.6, June 2014, pp. 4814 ~ 4824
DOI: 10.11591/telkomnika.v12i6.5541

A Dominance Degree for Rough Sets and Its Application in Ranking Popularity

Jia Zhao*[1], Jianfeng Guan[2], Changqiao Xu[2,3] and Hongke Zhang[1]
[1]National Engineering Laboratory for Next Generation Internet Interconnection Devices,
Beijing Jiaotong University, Beijing, China
[2]State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing, China
[3]Institute of Sensing Technology and Business,
Beijing University of Posts and Telecommunications, Wuxi, Jiangsu, China
*Corresponding author, e-mail: 11111004@bjtu.edu.cn

Abstract

Rough set theory is used in data mining through complex learning systems and uncertain information decision from artificial intelligence. For multiple attribute decision making, rough sets employ attribute reduction to generate decisive rules. However, dynamic information databases, which record attribute values changing with time, raise questions to rough set based multiple attribute reduction. This paper proposes a dynamic attribute based dominance degree for rough-set-based ranking decision. According to the dominance relations between two objects in dynamic information table, we propose three judgments and their judging values to get a dominance degree value, by which we can deny or approve of the dominance relations. Then we use the dominance-degree-based rough set to make dynamic attribute reduction. We applied this method in ranking popularity of network service resources. and extract ranking decision rules. Experiments show comparison between the searching engines with and without the ranking function and the efficiency of rough set ranking by our proposed dominance degree value.

Keywords: *rough set, judging value, dominance degree, dynamic attribute, popularity ranking*

1. Introduction

Developments in machine learning and data mining expand the dimension of capability in pattern recognition and enlarge the knowledge domain [1]. Diverse features or patterns are mined as well as some judging rules extracted for recognizing of new objects. Learning process can be regarded as either classifying or clustering process [2]. Supervised learning systems use introduction and classification model to identify objects, while unsupervised systems take some discriminatory judgments to cluster objects and explain their similarity. Pawlak [3] proposed rough set theory and attribute reduction to solve multi-attribute decision making problems in machine learning and artificial intelligence. Rough set, employed as unsupervised learning method, is suitable for uncertain information processing. Rough function was also proposed to describe the approximation to complex functions or patterns [4]. The concept and techniques are applied widely in machine learning and artificial intelligent system designing. Rough set make full use of binary relation between each two objects. Original rough set theory discusses the indiscriminate binary relation and use attribute information table to group objects [5]. There are also other binary relations. Slowinski et al. [6] proposed similarity relation based rough approximations. Greco [7] proposed dominance relation based rough set method to solve multiple attribute decision problems. In applications, rough set can either be used in decision making in data mining, or assist as pretreatment system with other learning and classifying methods such as neural network or fuzzy set. Dominance relation based rough set framework is employed in multiple attribute decision making when one decisive group is preferred to another one on an attribute in the information tables. Dominance degree describes how much one object outperforms another one about an attribute [8]. As an efficient soft computing method, rough set has been widely applied in the realm of information science. Li et al. [9] proposed a Rough sets-based search engine for grid service discovery. Rough set based multiple attribute decision have been used in airport service quality ranking [10, 11]. Dominance based rough

Received December 31, 2013; Revised March 16, 2014; Accepted March 29, 2014

sets approach has been applied in evaluating the impact of information technology. Kaneiwa et al. [12] used rough set based decision method to mine the sequential pattern. Factors affecting the adoption of Software service were analyzed with rough set decision method. Rough set plays its role to deal with static information tables.

However, there are so many dynamic database and dynamic information tables to record frequently changed and updated attribute values of an object [13]. Traditional static information tables do not reflect variation of a value over a period of time. Some complicated dynamic databases can record the variation sequences for an attribute item of objects in a ranking system [14]. For example, we analyze service popularity of a campus network in our experiments. In this example, some attributes of a content item are dynamically changed over an observation time. These attributes can be some kind of increasing statistics such as click rates or visiting times. How to deal with these dynamic attributes and embody variation in decision mechanism of popularity becomes a tough question, because popularity has to be recognized as a complex pattern, which deserves deliberate design of ranking method.

In this paper, we propose a dominance degree for rough set-based popularity ranking to solve dynamic attribute decision making problem and its application in a network service ranking and searching system. We give three dominance relation cases and propose judging values on each attribute in three cases: (i) in the basic dominance judgment, we judge whether object A outperform B in comparatively more observation times; (ii) in the frequently changing judgment, we judge whether A and B change their dominant relation frequently in different observation times; (iii) in the greatly changing judgment, we judge whether at some observation time points one object outperforms another one very much to a large enough degree. With these three degree values, we can calculate the dominance degree judging value. If the dominance degree value is greater than a threshold, we judge there is a dominance relation between two objects. If the dominance degree value is less than a threshold, we deny the dominance relation. After assign a dominance degree value to each attribute, we obtain all dominance relations between each two objects in information table. Then we use rough set-based attribute reduction to make a multiple attribute decision and employ these methods to rank the popularity of the campus network service resource. Experimental results show that the system with our proposed ranking machinery can efficiently react to the variation in dynamic attribute information table and outperform the situation without ranking. The main contributions of this paper includes: 1) We recognize popularity as a complex pattern that has multiple attributes and deserves elaborate learning method; 2) We give three case to describe the variation of attributes in dynamic information table and propose three judging value to obtain a dominance degree; 3) We apply this dominance degree in the rough set-based decision to extract rules for ranking popularity.

The remainder of this paper is organized as follows: Section II introduces some related theories including learning function and rough set. In section III, we propose the three judging values and the dominance degree for rough set. In addition, we prove some properties of the dominance degree based on rough set. In Section IV, we do the experiment based on a network service retrieval system. Section V concludes the paper.

2. Research Method

2.1. Degree Function for Dynamic Attributes

Dynamic attribute data always exits in information process. Table 1 shows click rates of two web videos in seven days. Click rates stands for attribute of popularity in some way. These popularity values are variable during a week. Considering that more popular video is preferred, we need to decide whether there is a dominance relation between two videos. New method should be introduced to ranking decision by using an attribute whose value changes frequently over an observation time. On the other hand, although there is not an obvious great or little numerical relation between two dynamic attribute sequences, we can use a degree function extracted from attribute performing statistics over a time to analyze the dominance relation. Let A and B are two objects with a dynamic attribute. We can see performance of the attribute as two sequences S_A and S_B, in which a_n and b_n are values in time point n.

$$S_A = \{a_1, a_2, ...a_n, ...\} \quad (1)$$

$$S_B = \{b_1, b_2, ...b_n, ...\} \quad (2)$$

Each couple of (a_n, b_n) has a dominating or dominated relation. For finding way to decide the preferred relation between the two whole sequences, we use these value couples to figure out a statistic result. Numerical relation between a value couples can be express as a point in a coordinate system where dominance relation means the value couple point above the line $y = x$, and dominated relation means the value couple point below the line $y = x$.

Table 1. Click Rates of Web Video A an B in Seven Days

ID	Click rates						
	Day1	*Day2*	*Day3*	*Day4*	*Day5*	*Day6*	*Day7*
A	96	215	299	411	658	872	1305
B	118	187	351	564	643	945	1138

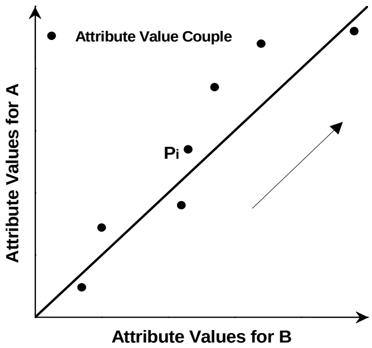

Figure 1. Basic Dominance Judgement

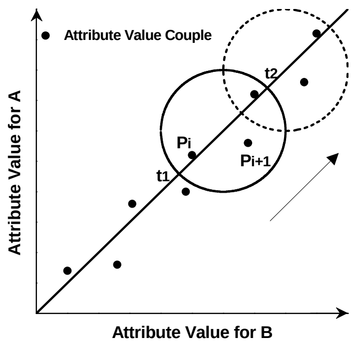

Figure 2. Frequently changing judgement

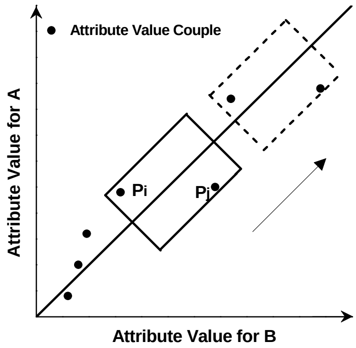

Figure 3. Greatly changing judgement

2.2. Dominance Judging Values

Consider three cases: (i) One object has more values greater than another in the dynamic attribute sequence. (ii) Preference relation between two objects change frequently over a time. (iii) One object has a value much greater than another object in the same position of the dynamic attribute sequence. Focusing on the three cases, we propose three judging values to calculate the dominance degree value, by which we approve or deny the dominance relation between each two objects.

2.2.1. Basic judging value

In Equations (1) and (2), we express the dynamic attribute as two sequences and give the value couple in same position of the sequence as (a_n, b_n). To analyze the dominance relation couple in coordinate system, we use the point (x_i, y_i) to present the value couple in position i of the sequence. Length of sequence can be defined as N. So there are N points in the coordinate. As shown in Figure 1, the point above the line $y = x$ indicates that on the dynamic attribute object A has a value preferred to object B in the same position of the sequence. Then we calculate the total number of points above the line $y = x$ and define this number as n. So the probabilistic value of A dominating B can be expressed as follow:

$$p_b = \frac{n}{N} \tag{3}$$

Considering the dominance degree in dominance relation, when degree value greater than zero, dominance relation exists. Otherwise, a dominated relation exists. We can use the dominance probabilistic value to act as plus or minus sign in dominance relation judgment. Given a threshold value $p_{b_T} = 0.5$, we temporarily approve of the dominance relation and present it with plus degree d^+ on the condition that our basic judging value is greater than p_{b_T}. Correspondingly, when p_b is less than p_{b_T}, we directly deny the dominance relation and present it as minus degree d^-. The basic judging value p_b also contributes to the dominance degree value *h*. Because *h* eventually decides the dominance relation, basic judging value is a temporary judgment and the first step to judge a dominance relation.

2.2.2. Frequently Changing Judging Value

When one object satisfies the condition $p_b \geq p_{b_T}$ to another object, whether a dominance relation exists is still a question. Because there are other situations about distribution of attribute value couples, we cannot directly judge a dominance relation just by more points above the line $y = x$. let us introduce another situation in which we can also get a judging value and make it contribute to the final dominance degree value *h*.

As shown in Figure 2, two neighbor points P_i and P_{i+1} respectively stay on one side of the line $y = x$. If we connect the two neighbor points with a line, there will be cross point with the line $y = x$. We do the same thing to each two neighbor points across the line $y = x$. Then we calculate the total number of the cross points on the line $y = x$. To accomplish this, our first step is to judge whether there is a cross point between the neighbor points. We use the vector triangle area method to judge whether the two neighbor points are separated on both sides of the line $y = x$. considering the neighbor points P_i and P_{i+1} in the geometric figure, we choice any two points, named t_1 and t_2, on the line $y = x$. Coordinates of P_i and P_{i+1} are (x_i, y_i) and (x_{i+1}, y_{i+1}), while coordinate of t_1 and t_2 are (a_1, b_1) and (a_2, b_2). We connect the point couples to obtain some vectors $\overrightarrow{t_1t_2}$, $\overrightarrow{t_2P_i}$, $\overrightarrow{P_it_1}$, $\overrightarrow{t_2P_{i+1}}$ and $\overrightarrow{P_{i+1}t_1}$. Then we calculate areas of the two vector triangles as follows:

$$S_i = \begin{vmatrix} a_1 & a_2 & x_i \\ b_1 & b_2 & y_i \\ 1 & 1 & 1 \end{vmatrix} \quad S_{i+1} = \begin{vmatrix} a_1 & a_2 & x_{i+1} \\ b_1 & b_2 & y_{i+1} \\ 1 & 1 & 1 \end{vmatrix} \tag{4}$$

If $S_i \cdot S_{i+1} < 0$, there is a cross point on the line $y = x$. If $S_i \cdot S_{i+1} > 0$, there is no cross point between two neighbor points. Next step is to calculate the total number of the cross points on the line $y = x$. Because the dynamic attribute has an increasing sequences and every position in the sequence is a sampling time point over a period of time, we regard the line $y = x$ in the geometric figure as a time axis. The attribute value points are distributed near the line in the increasing direction across the line. At any time point t on $y = x$, we use a circle with radius *f* as an observation window, which can slide across the time line $y = x$. We calculate the number of cross points in the window at time *t* with the following equation:

$$n_f = \sum_{i=1}^{n} I_{fi}(t) \tag{5}$$

Where the number of attribute points in the window is *n*, and the variable $I_{fi}(t)$ is defined as below:

$$I_{fi}(t) = \begin{cases} 1 & if \quad S_i \cdot S_{i+1} < 0 \\ 0 & if \quad S_i \cdot S_{i+1} > 0 \end{cases} \tag{6}$$

Here we discuss the time points. In applications such as network resource popularity ranking, attribute values of recent period of time are more referentially important. Accordingly, we add a weight $w_f(t)$ to each time point. The weight function is increasing and satisfies the equation $\int_0^T w_f(t) = 1$. Here T is the period of observation time. To normalize the judging values, we use the ratio of cross point amount and attribute point amount $N_f(t)$ in the window to express the judging value at time t as bellow:

$$p_f(t) = w_f(t) \cdot \frac{\sum_{i=1}^{n} I_{fi}(t)}{N_f(t)} \tag{7}$$

As the window across the time line $y = x$ over a period of time T, we can express the frequently changing judging value as follow:

$$p_f = \frac{1}{T}\int_0^T p_f(t)\mathrm{d}t \tag{8}$$

2.2.3. Greatly Changing Judging Value

In this part, we discuss the situation in which some attribute points are far from the time line *y=x* in a great distance. We analyze the features of attribute points both above and below the time axis *y=x*. One of the most important feature is the distance from attribute point to the line *y=x*. This feature somehow presents the dominance extent on one attribute point. In this situation, we use a rectangle window across the time line to observing the attribute points. Half of the window length is greater than the maxim distance from a point to the line *y=x*.

In the window at time t, we can find a max-distance point on both side of time line *y=x*. we express the judging value at time *t* as the ratio of max-distance on one side to the sum of the max-distances on both sides. As shown in Figure 3, in the window at time *t*, $P_i = (x_i, y_i)$ represents the max-distance point above the time axis, while $P_j = (x_j, y_j)$ represents the max-distance point below the time axis. The judging value at time t is defined as follow:

$$p_g(t) = w_g(t)\frac{|x_i - y_i|}{|x_i - y_i| + |x_j - y_j|} \tag{9}$$

Where $w_g(t)$ is an increasing weight function whose value larger across the timeline and satisfying $\int_0^T w_g(t) = 1$. As the rectangle window across the time axis over a period of time *T*, we can obtain the greatly changing judging value below:

$$p_g = \frac{1}{T}\int_0^T p_g(t)\mathrm{d}t \tag{10}$$

2.2.4. Dominance Degree Value

To summarize the cases above, we obtain three judging values about the dominance relation between two dynamic attribute sequences. Three judging values all contribute to the dominance degree value *d*. and the basic judging value is also used in the judging of sign. We defined the dominance degree value as follow:

$$d = \mathrm{sgn}(p_b > 0.5) \cdot p_b \cdot p_f \cdot p_g \tag{11}$$

We also give a dominance threshold d_T. When the dominance degree value is greater than d_T, we approve of the dominance relation. Correspondingly, we can deny the dominance relation if the dominance degree is less than d_T.

2.3. Rough Sets with Dynamic Attributes

There is a corresponding binary relation between two objects A and B in rough set. The relation of A dominating B can be regard as the relation of B dominated by A. The dominance degree threshold d_T should be expressed in two formats of d_T^+ and d_T^-.

We use $yP^d x$ to presents the dominance relation. P is a dynamic attribute set. d is the dominance degree value. Object x and y all belong to the object set U. The dominance set, whose members all dominate object x, is defined as follows:

Definition 1. On the dynamic attribute set P, Given an object x, for objects $y, y' \in U$, if $yP^{d'}x$, we define the set $D_P^d(x) = \{y \in U : yP^{d'}x, d' \geq d > 0\}$ as the dominance set, and if $y'P^{d'}x$, then we define the set $D_P^d(x) = \{y' \in U : y'P^d x, d' \leq d < 0\}$ as the dominated set. According to the definition above, we give two properties about the dominance set and the dominated set.

Property 1. On the dynamic attribute set P, for the threshold d_T^+, two dominance degree values $d_1, d_2 \in [d_T^+, 1]$ and the object $x \in U$, if $d_1 \geq d_2 \geq d_T^+ > 0$, then the dominance sets satisfy $D_P^{d_1}(x) \subseteq D_P^{d_2}(x)$.

Property 2. On the dynamic attribute set P, for the threshold d_T^-, two dominance degree values $d_1, d_2 \in [-1, d_T^-]$ and the object $x \in U$, if $d_1 \leq d_2 \leq d_T^- < 0$, then the dominated sets satisfy $D_P^{d_1}(x) \subseteq D_P^{d_2}(x)$.

Definition 2. On the dynamic attribute set P, for the object set U, dominance degree value d and the decisive set X, the up and down approximations are defined as follows:

$$\underline{P}(X) = \bigcup_{d \in [d_T^+, 1]} \{y \in U : x \in U, y \in D_P^d(x), D_P^d(x) \subseteq X\}, \quad \overline{P}(X) = \bigcap_{d \in [d_T^+, 1]} \{y \in U : x \in U, y \in D_P^d(x), X \subseteq D_P^d(x)\}$$

Definition 3. On the dynamic attribute set P, for the object set U, dominance degree value d and the decisive complement set X^C, the up and down approximations are defined as follows:

$$\underline{P}(X^C) = \bigcup_{d \in [-1, d_T^-]} \{y \in U : x \in U, y \in D_P^d(x), D_P^d(x) \subseteq X^C\}$$

$$\overline{P}(X^C) = \bigcap_{d \in [-1, d_T^-]} \{y \in U : x \in U, y \in D_P^d(x), X \subseteq D_P^d(x)\}$$

Theorem 1. On the dynamic attribute set P, for the decisive set X, if the dominance degree value $d \in [d_T^+, 1]$, then the up and down approximations can be expressed as follows:

$$\underline{P}(X) = \{y \in U : x \in U, y \in D_P^{d_T^+}(x), D_P^{d_T^+}(x) \subseteq X\}, \quad \overline{P}(X) = \{y \in U : x \in U, y \in D_P^1(x), X \subseteq D_P^1(x)\}$$

Proof . The dominance degree value satisfies the condition $d \in [d_T^+, 1]$. We can get that that any degree value d_i is not less than the threshold value d_T^+. According to the property 1, from the relation $d_i \geq d_T^+$, we can get the relation between the two dominance sets as $D_P^{d_i}(x) \subseteq D_P^{d_T^+}(x)$. For all possible d_i, we get $\bigcup_{d_i \in [d_T^+, 1]} D_P^{d_i}(x) \subseteq \bigcup_{d_i \in [d_T^+, 1]} D_P^{d_T^+}(x) = D_P^{d_T^+}(x)$.

Also, the belonging relation $d_T^+ \in [d_T^+, 1]$ leads to the including relation $D_P^{d_T^+}(x) \subseteq \bigcup_{d \in [d_T^+, 1]} D_P^{d_i}(x)$. Then we get the relation $D_P^{d_T^+}(x) = \bigcup_{d \in [d_T^+, 1]} D_P^{d_i}(x)$. For the definition of the up approximation, if the union $\bigcup_{d \in [d_T^+, 1]} D_P^{d_i}(x) \subseteq X$, then $D_P^{d_T^+}(x) \subseteq X$. Therefore, the up approximation is expressed as theorem 1.

According to the property 1, when $d_T^+ \le d_i \le 1$, the relation between the two dominance sets is $D_P^1(x) \subseteq D_P^{d_i}(x)$. We can get the relation below: $\bigcap_{d_i \in [d_T^+,1]} D_P^1(x) \subseteq \bigcap_{d_i \in [d_T^+,1]} D_P^{d_i}(x) \subseteq D_P^1(x)$

The dominance sets satisfies $\bigcap_{d_i \in [d_T^+,1]} D_P^{d_i}(x) = D_P^1(x)$. if $X \subseteq \bigcap_{d_i \in [d_T^+,1]} D_P^{d_i}(x)$, then $X \subseteq D_P^1(x)$. So the down approximation is expressed as theorem 1.

Theorem 2. On the dynamic attribute set P, for the decisive complement set X^C, if the dominance degree value $d \in [-1, d_T^-]$, then the up and down approximations are as follows:

$\underline{P}(X^C) = \{y \in U : x \in U, y \in D_P^{d_T^-}(x), D_P^{d_T^-}(x) \subseteq X^C\}$, $\overline{P}(X^C) = \{y \in U : x \in U, y \in D_P^{-1}(x), X^C \subseteq D_P^{-1}(x)\}$

Proof. For any possible dominance degree value $d_i \in [-1, d_T^-]$, according to property 2 and the including and belonging properties in sets, we can get the relations below:

$$\bigcup_{d_i \in [-1, d_T^-]} D_P^{d_i}(x) \subseteq \bigcup_{d_i \in [-1, d_T^-]} D_P^{d_T^-}(x) = D_P^{d_T^-}(x) \quad , \quad D_P^{d_T^-}(x) \subseteq \bigcup_{d_i \in [-1, d_T^-]} D_P^{d_i}(x)$$

Then the union of dominance sets satisfies the relation $D_P^{d_T^-}(x) = \bigcup_{d_i \in [-1, d_T^-]} D_P^{d_i}(x)$. if $\bigcup_{d_i \in [-1, d_T^-]} D_P^{d_i}(x) \subseteq X^C$, then $D_P^{d_T^-}(x) \subseteq X^C$. Therefore, the up approximation follows as theorem 2.

For $-1 \le d_i \le d_T^-$, we can get the relation: $\bigcap_{d_i \in [-1, d_T^-]} D_P^{-1}(x) \subseteq \bigcap_{d_i \in [-1, d_T^-]} D_P^{d_i}(x) \subseteq D_P^{-1}(x)$

The dominance sets satisfies $\bigcap_{d_i \in [-1, d_T^-]} D_P^{d_i}(x) = D_P^{-1}(x)$. If $X^C \subseteq \bigcap_{d_i \in [-1, d_T^-]} D_P^{d_i}(x)$, then $X^C \subseteq D_P^{-1}(x)$. So the down approximation is expressed as theorem 2.

Theorem 3. The object set U can be grouped into two decisive set X and its complement X^C. When a new object y joins the set U and need to be classified, for the objects $x \in X$ and $x' \in X^C$, with the dominance degree value d of y, if $d > d_T^+$, then $y \in X$; if $d < d_T^-$, then $y \in X^C$.

Proof. Considering the threshold d_T^+ and the object $x \in X$, we can see that if $D_P^{d_T^+}(x) \not\subset X$, then $D_P^{d_T^+}(x) \cap X^C \ne \varnothing$. But the property that objects in set X^C are dominated by x contradicts the definition of $D_P^{d_T^+}(x)$. Therefore, if $d > d_T^+$, according to the property 1, we get $y \in D_P^d(x) \subseteq D_P^{d_T^+}(x) \subseteq X$.

For the threshold d_T^- and the object $x' \in X^C$, if $D_P^{d_T^-}(x') \not\subset X^C$,then $D_P^{d_T^-}(x') \cap X \ne \varnothing$. The property that objects in X dominate x' contradicts the definition of $D_P^{d_T^-}(x')$. Therefore, if $d < d_T^-$, according to the property 2, we can get the relation: $y \in D_P^d(x') \subseteq D_P^{d_T^-}(x') \subseteq X^C$

Theorem 4. *On the dynamic attribute set P, for two objects* $x, y \in U$ *, in the relations* $yP^{d^+}x$ *and* $xP^{d^-}y$ *, where* $d^+ > 0$ *and* $d^- < 0$*, if* $d^+ = d_{\max}^+$*, then* $d^- = d_{\min}^-$.

Proof. Considering the equation (11), we can see $p_b, p_f, p_g > 0$. Given the constant p_f, we can get: $\left|d^+\right| = p_f \cdot p_b \cdot p_g \le p_f \cdot \left(\frac{p_b + p_g}{2}\right)^2$, $\left|d^-\right| = p_f \cdot (1 - p_b) \cdot (1 - p_g) \le p_f \cdot \left(\frac{2 - p_b - p_g}{2}\right)^2$.

If $d^+ = d_{\max}^+$, then $p_b = p_g$. When $p_b = p_g$, $d^- = d_{\min}^-$.

3. Results and Discussion

In this section, we will evaluate the performance of our proposed dynamic attribute decision making method. Firstly, we give a simple example to illustrate rough set based ranking decision with dynamic attributes. Then, we employ our proposed method in network service resource ranking with our campus networking flow statistics. Finally we exam efficiency of a network resource searching engine implemented with rough set based dynamic attribute ranking decision.

Table 2. Information Table About Web Videos

Web video	Duration	Flow	Visiting time in 5 days	rank
x_1	13	124.5	{467, 771, 1109, 1415, 2007}	1
x_2	19	158.47	{436, 864, 1051, 1662, 1932}	2
x_3	11	76.9	{89, 201, 227, 453, 539}	2
x_4	17	58.4	{43, 89, 415, 447, 486}	2
x_5	22	160.3	{225, 423, 671, 849, 1070}	1

Table 3. Dominance Judging Results

Relation couple	P_f	P_b	P_g	d	Dominance
(x_1, x_2)	0.2	0.6	0.223	0.028	No
(x_1, x_3)	1	1	1	1	Yes
(x_1, x_4)	1	1	1	1	Yes
(x_1, x_5)	1	1	1	1	Yes
(x_2, x_3)	1	1	1	1	Yes
(x_2, x_4)	1	1	1	1	Yes
(x_2, x_5)	1	1	1	1	Yes
(x_3, x_4)	0.6	0.8	0.373	0.179	No
(x_5, x_3)	1	1	1	1	Yes
(x_5, x_4)	1	1	1	1	Yes

Web videos have some flow statistic features that can be used as dynamic attributes for their popularity ranking decision. To clarify the ranking decision process, let's work through a web video ranking example containing three dominance relation attributes, one of which is a dynamic attribute. Table 2 shows ranking decision information of the five web videos. There are three statistic attribute containing duration, flow and visiting times in five days. The first column of the table displays the video names listed in order from x_1 to x_5. In second column of the table, visiting duration is displayed with time unit of hour. The third column shows flow data for each video with the unit of GB. Both this two attributes embody dominance relation between each two web videos and show popularity expressed by this preferring relation. The values of this dynamic attribute for each video are shown in the format of sequence. There are also some features that we can see from the fourth column. Relation of value sequences between two videos also presents preference. And the attribute value sequences are increasing sequences. It is easy to understand that the popular attribute like visiting times is an accumulative variable whose value always increases as more and more people click and visit the resource. The fifth column displays the decisive rankings. According to the attributes and decisive ranking groups we can process this information table and do the multiple attribute decision making with rough set based dynamic attribute ranking. Table 3 shows the analysis of the dominance relations. The first column lists the dominance relation couples. The second column displays the value of frequently changing judgment variable for each dominance relation couple. Values of basic judgment variable calculated by Equation (3) for each dominance relation couple are show in the third column. The fourth column shows the values of greatly changing judgment variable. The fifth column lists the dominance degree values for each dominance relation couple.

According to dominance degree values in fifth column, dominance decisions are shown in the last column. Notice the second row in the table, the dominance degree value h is little than the threshold value d_T , which is given by d_T =0.3. So, the dominance relation between x_1 and x_2 on the attribute of visiting times is denied. In the same way, the dominance relation between x_3 and x_4 on this dynamic attribute are denied. You can see all the other couples have a degree value equaling to 1, and these couples definitely have a dominance relation. From the data of the table, we can see that too little degree value will be used to deny a dominance relation although one object has more dominance points than another.

Table 4. Ranking Rules from Discriminate Matrix

Rank 1 to rank 2	x_2	x_3	x_4	Dominating Rules
x_1	dfc	fc	Ø	f≥124.5, c≥c(x_1)
x_5	dfc	dfc	dfc	d≥22, f≥160.3, c≥c(x_5)
Dominated Rules	d≤19 f≤158.47 c≤c(x_2)	d≤11 f≤76.9 c≤c(x_3)	f≤58.4 c≤c(x_4)	

The next step of this example is to make a multiple attribute decision by some decision rules. As shown in Table 4, the web videos are display in two groups by their ranking value. Firstly, in each position of the row corresponding to the dominated group x_2, x_3 and x_4, there is an attribute variable name character string, each character of which express that the attribute can discriminate the two objects by definite dominance relation. According to the data in Table 2 and Table 3, we can fill this attribute character string in the position related with a dominance relation couple. Then, from this table, we can extract ranking decision rules by using the discriminate function.

We record the network traffic statistics over a period of time and analyze many network flow features that can be used to describe and decide the popularity. We collected two month statistics, totally 425.73 TB upload and download flows and 19907944 times resource views. 500 registered services are classified into web page, video, images and others. We obtain totally 16 kinds of flow attributes from our statistics. Our goal is to make a multiple attribute and multi-relation based rough set ranking decision. In these attributes, category and size are indiscriminate relation based static attributes. Content and protocol are similarity relation based static attributes. There are 9 dominance-relation-based dynamic attributes: access time, daily views, daily visitors, key words, comments, flow, duration, resource view and unique visitor.

Figure 4 shows the comparison of the ranking results with and without the dynamic attribute reduction between two resources. The basic judging value is greater than 0.5, but the dominance degree value is less than the threshold. Each resource is ranked by percentage number of resources below it. Results of Figure 4(a) and (b) lie in the too little frequently changing judging value, which leads to the dominance degree value less than the threshold. Without this judging value, one resource is classified to a high rank set while the other resource is grouped into a low rank set. With this judging value, dominance relation between the two resources is denied. Therefore, two resources are ranked nearly. As to effect of the greatly changing judging value shown in Figure 4(c) and (d), it is analogous to the analyzing of frequently changing judging value above.

Figure 5 shows the comparison of searching result relevance with and without ranking function in 10 queries. We use this result to evaluate the efficiency of resource ranking in searching. We evaluate the searching result items with the Normal Discounted Cumulative Gain (NDCG) value, which is widely used in evaluating the performance of searching engines [18]. We can see the searching engine with the popularity ranking function show higher value of NDCG than just key word searching without resource ranking. Explanation is that among all the key-word matched results, more popular resources have a high possibility to satisfy users.

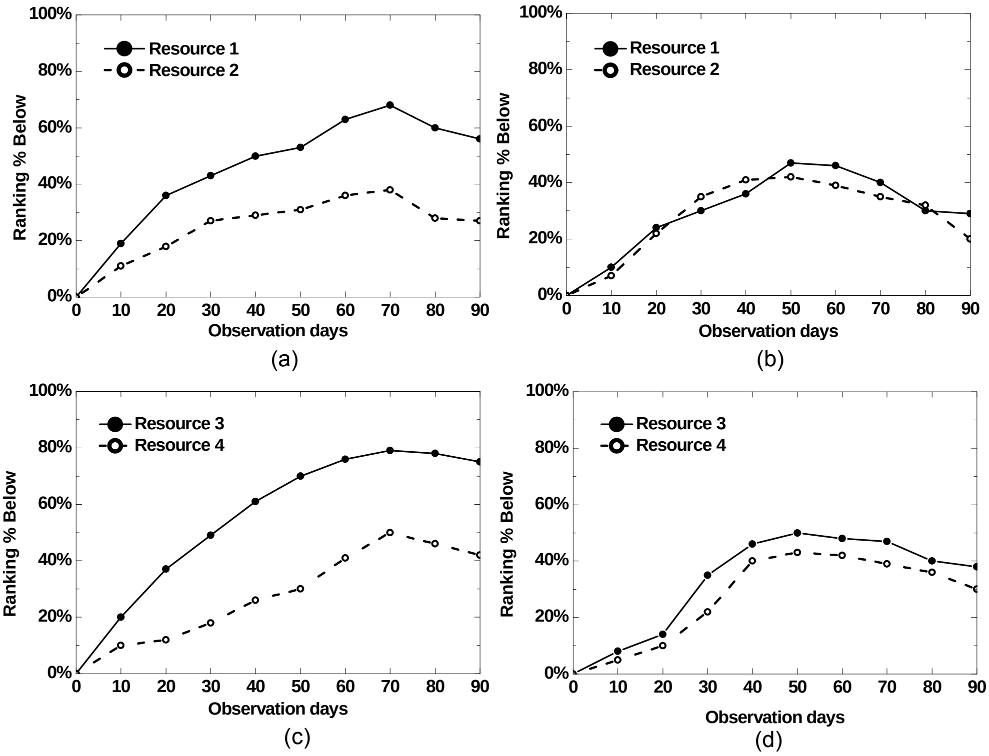

Figure 4. Comparison of Two Resource Ranking with and without Dominance Judging Values

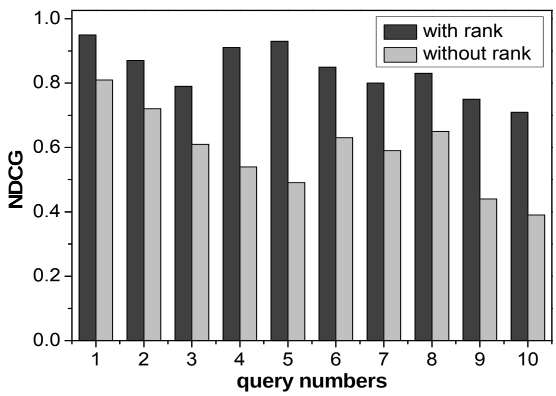

Figure 5. Comparison of NDCG in Web Searching with and without Popularity Ranking

4. Conclusion

In this paper, we propose a dynamic attribute based dominance degree value for rough set ranking decision. Dominance relation between two objects may be decided by a dynamic attribute whose value is not a single number but a sequence. Items of the sequence are the sampling values at different time over an observation time. To solve the problem on dominance relation judging by dynamic attributes, we propose three judging values respectively in three geometric cases about distribution of attribute value couple points. With the three judging values, we obtain the dominance degree value, by which we can decide the dominance relation. Then we build the dominance relation based rough sets with the dynamic attributes and dominance degree values. The most important application of dominance relation based rough set lies in multiple attribute ranking decision. We use rough set based dynamic attribute reduction by the dominance degree value to rank the campus network service resources.

Experiments show comparison between the searching engines with and without the ranking function and the efficiency of rough set ranking by our proposed dominance degree value.

Acknowledgements

This work is supported by the National Basic Research Program of China (973 Program) under Grant 2013CB329101, and partially supported by the National High-Tech Research and Development Program of China (863) under Grant No. 2011AA010701, and by the National Natural Science Foundation of China (NSFC) under Grant No. 61232017, 61003283, 61001122, 61372112, and by the Jiangsu Natural Science Foundation of China under Grant No. BK2011171, and by the DNSLAB project (China Internet Network Information Center) under contact No.DNSLAB-2012-N-U-13.

References

[1] RW Swiniarski, A Skowron. Rough Set Methods in Feature Selection and Recognition. *Pattern Recognition Letters*. 2003; 24(6): 833-849.
[2] YY Yao. Information Granulation and Rough Set Approximation. *International Journal of Intelligent Systems*. 2001; 16(1): 87-104.
[3] Z Pawlak. Rough Sets. *International Journal of Information and Computer Sciences*. 1982; 11(5): 341-356.
[4] Z Pawlak. Rough Sets. Rough Relations and Rough Functions. *Fundamenta Informaticae*. 1996; 27(2-3): 103-108.
[5] Z Pawlak, A Skowron. Rudiments of Rough Sets. *Information Sciences*. 2007; 177(1): 3-27.
[6] R Slowinski, D Vanderpooten. A Generalized Definition of Rough Approximations based on Similarity. *IEEE Trans. Knowledge and Data Engineering*. 2000; 12(2): 331-336.
[7] S Greco, B Matarazzo, R Slowinski. Rough Approximation of a Preference Relation by Dominance Relations. *European Journal of Operational Research*. 1999; 117(1): 63-83.
[8] S Greco, B Matarazzo, R Slowinski. Rough Sets Theory for Multicriteria Decision Analysis. *European Journal of Operational Research*. 2001; 129(1): 1-47.
[9] MZ Li, B Yu, O Rana, ZD Wang. Grid Service Discovery with Rough Sets. *IEEE Trans. Knowledge and Data Engineering*. 2008; 20(6): 851-862.
[10] JJH Liou, CH Tang, WC Yeh, CY Tsai. A Decision Rules Approach for Improvement of Airport Service Quality. *Expert Systems with Applications*. 2011; 38(11): 13723–13730.
[11] JJH Liou. Variable Consistency Dominance-based Rough Set Approach to formulate airline service strategies. *Applied Soft Computing*. 2011; 11(5): 4011–4020.
[12] K Kaneiwaa, Y Kudob. A Sequential Pattern Mining Algorithm using Rough Set Theory. *International Journal of Approximate Reasoning*. 2011; 52(6): 881–893.
[13] CF Ahmed, SK Tanbeer, B Soo Jeong, YK Lee. Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases. *IEEE Trans. Knowledge and Data Engineering*. 2009; 21(12): 1708-1721.
[14] KT Chuang, KP Lin, MS Chen. Quality-Aware Sampling and Its Applications in Incremental Data Mining. *IEEE Trans. Knowledge and Data Engineering.* 2007; 19(4): 468-484.