Application of enhanced rough set approach to the evaluation of urban water resources utilization: A case study of Beijing

Lvxiang Deng [a,b]
[a] School of Geo-science and Environmental Engineering, Central South University (CSU),Changsha, China,
[b] Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada
lxdeng@mail.csu.edu.cn

Bryan Karney
Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada
karney@ecf.utoronto.ca

Abstract— **Comprehensive evaluation of urban water utilization, which has interested researchers for years, is a pre-requisite for optimal urban water resources management. Moreover, an accurate assessment of urban water use can greatly assist municipal managers and policy makers. This article presents an enhanced rough set method for generating evaluation rules from historical local data. The proposed method extends classical rough set theory by using a conditional probability classifier for equivalence class and decision rules to deal with any incomplete and ambiguous undecided objects. The approach is tested using data from Beijing whose water scarcity problem not only represents a microcosm of China but generally illustrates the value of the new approach.**

Keywords-Urban water utilization; knowledge discovery; rough set; comprehensive evaluation

I. INTRODUCTION

A large proportion of the world's population has recently been affected by water scarcity, especially in areas with high population density, low average precipitation and under-developed economies. Indeed, water is often a key regional restricting factor on economic growth. The growing difficulty of exploiting new water sources, increasing environmental and labour costs, and the already high rates of water utilization have all redirected attention from the traditional preoccupation with water-supply-oriented management to improved water use and better water demand management. How to evaluate and limit the urban water use in the light of environmental constraints has become a crucial research question. Especially recently, the evaluation of water resources use – including the evaluation of local water systems, water deployment, local water carrying capacity[1], water quality[2], water exploitation[3] and water resources renewability[4,5] – have become pressing issues especially in regions with severe water scarcity such as China. Many methods of analysis had been applied, including, Analytic Hierarchy Process (AHP), Delphi, Fuzzy theory, Artificial Neural Networks (ANN) and Grey Clustering Analysis.

Three steps are frequently used to estimate the water status: providing suitable indicators, giving the weight of each indicator and then calculating certain composite values which together function as a "grade" of the water system. The shortcomings of such an approach are evident. Certainly, the key issue of evaluation an area's water use status is to how to assign a reliable and reasonable weight to each indicator. But there are no objective methods of doing this. The resulting assessment is currently poorly understood even by experts who inevitably approximate water utilization status based on personal experience. The often inaccurate estimations might result in inefficient and inappropriate strategies for either management or further development.

An alternative approach for assessment is that discovering rules from the existed information automatically, effectively constructing or refining a training sample set from observed data and through an algorithm. One appropriate was of doing this uses the Rough Set approach which to date has had only a few applications such as evaluating water renewability[6], assessing local water systems[7] or predicting water demand[8]. Rough Set theory was proposed by Pawlak in 1982[9]; in general, it is a mathematical tool with many applications in decision analysis and support. Its basic philosophy is making use of on object set to describe the knowledge and discover rules from the given set.

This article provides a brief overview of the current research, and constructs a comprehensive indicator system that covers all necessary factors of water utilization, environment and socio-economy based on concepts of "water value", "sustainable development", and a review of the local ecosystem. Then we present an enhanced rough set method for generating rules from a set of historical local observed data to assess water utilization status. The proposed model attempts to reduce the dependence on experts' opinions of the former evaluation methods and to extend the standard rough set method by using a conditional probability classifier of equivalence class and decision rules to deal with the incomplete and ambiguous undecided objects. Finally, Beijing, whose water scarcity problem represents a microcosm of China, is used to illustrate the range experimental result as well as their possible value to decision and policy makers.

II. INDICATOR SYSTEM AND DATA REPRESENTATION METHOD

A. Indicator system of water utilization evaluation

For any urban area to operate normally, an enormous quantity of fresh water is typically consumed by industrial, commercial, public and domestic consumers; moreover, water distribution systems typically leak. With respect to urban development, including local economic growth, social development and natural protection, the sustainability of water resources use should take priority. In constructing an evaluation indicator system, there are four aspects that should be considered: water use efficiency, water security, water deployment and water environment, which thus imply the four evaluation principles of Efficiency, Security, Equity and Sustainability (ESES). Thus, twelve of many factors that may crucially affect the estimation of urban water utilization status are identified in an indicator system (see Table 1). Extensions of this list are obviously possible.

TABLE 1 INDICATOR SYSTEM OF ADMINISTRATIVE LEVELS

Object Level	Principle Level	Indicator Level
Comprehen-sive Evaluation of Urban Water Resources Sustainable Utilization	Water use Efficiency	a_1 Rate of water reuse in industry
		a_2 Water consumption per capita
		a_3 Water consumption for 10^4RMB GDP
	Water Security	a_4 Rate of water shortage
		a_5 Rate of contamination
		a_6 Degree of drought-flood disaster
	Water Deployment	a_7 Available water amount per capita
		a_8 Index of water supply
		a_9 Index of water deployment composition
	Water Environment	a_{10} Annual precipitation
		a_{11} Index of surface humidity
		a_{12} Rate of centralized sewage treatment

B. Data representation

Table 1 summarizes a set making up of condition factors and decision factors to which can be added a factor d, an assessment of the status of urban water utilization.

It is assumed that the above given set represents the knowledge about the domain. From the view of rough set theory, the set called *decision information system*, or referred to as a *classification system*. Before making use of rough set theory, the value range of all attributes should be provided.

For simplicity, these attributes are classified into three groups, (a) a_6, a_{10}, d; (b) a_1, a_4, a_5 ,a_{12}; (c) a_2, a_3, a_7, a_8, a_9, a_{11}. The values of attributes in the first group would be given by an experts group or obtained from existing researches. For example, a_6 (the Degree of drought-flood disaster in observed years) has a value chosen from the set {Very serious, Serious, Normal, Low, Inconsequential}. The a_{10}, the Annual Precipitation, is classified {Very low, Low, Normal, High, Very high}, d, the Decision attributes, {Very poor, Poor, Normal, Good, Very good}. The corresponding range prescribed by integer number is {1, 2, 3, 4, 5}.

For the second group, a_1, a_4, a_5 and a_{12}, all concern ratios or proportions. Given the current technology and compared with the developed countries, the rate of water reuse in industry and water treatment should be relatively high. So the value of these two factors should be subdivided as {≤40, 40-60, 60-80, 80-90, ≥90} corresponding to the integer number range {1,2,3,4,5}. The value of a_4, the rate of water shortage, is the ratio of the difference between the available water and the demand and the water demand amount. Indicator a_5 concerns about the fraction of polluted water in local water environment. It will be often be calculated by the proportion of the river that is contaminated. Hence, the value of a_4 and a_5 will be subdivided as {≥40, 30-40, 20-30, 10-20, ≤10}, corresponding to the integer number range {1,2,3,4,5}.

The finally group contains six attributes, a_2, a_3, a_7, a_8, a_9, a_{11}. They are obtained by the formula calculation from observed data. About a_2, a_3 and a_7 consider annual water use respectively in water consumption per capita (m^3/per person), Water consumption for 10^4RMB GDP ($m^3/10^4$ RMB GDP), Available water amount per capita (m^3/per person). Their value might be subdivided as {≤100,100-200,200-300, 300-400,≥400}, corresponding to the integer number range {1,2,3, 4,5}. The value of a_8, the index of the water supply, is considered the ratio of the amount of local water supply and ground area ($10^4 m^3/km^2$) and the range is {≤10, 10-15, 15-20, 20-30, ≥30}, the corresponding integer number range is {1,2,3,4,5}.

Indicator a_9 aims to consider the evolution of the structure of the water system from water consumption for each industry and their relationship by information entropy. In fact, water systems are open, exchanging information and energy with other systems. If the structure of water system is reasonable and stable, the system will tend perform well. Considering total water amount is Q, and there are N industries consume water. For each industry, the amount of water consumption is q_j, (j=0,1,…N). And $\sum_{j=1}^{N} q_j = Q$. p_j means the relative contribution of each industry and $p_j=q_j/Q$. Thus, the balanced information entropy could be calculated by $H = (-\sum_{j=1}^{N} p_j \ln p_j) / \ln N$ [10]. H_m=lnN means the maximum infomation entropy of this system. From the view of information entropy, the more unstable the system is, the more out-of-order it is, and the bigger the H value is. If the value of H does not change dramatically, the system is relatively stable and is not getting dramatically worse or better. Its range could be subdivided as {≤0.4, 0.4-0.6, 0.6-0.8, 0.8-0.9,≥0. 9} and the correspond integer number range is {1,2,3,4,5}.

Parameter a_{11}, the index of surface humidity, takes into account the effect of local weather. It's well-known that the weather changes and this is expressed by local precipitation and temperature change. The annual index of surface humidity might be calculated by $W = O / \sum_{i=1}^{12} O_{ei}$ [11]. O is the annual precipitation value (millimeter), O_{ei} means the potential evaporation amount of the i^{th} month and $\sum_{i=1}^{12} O_{ei}$ means annual potential evaporation amount. It might be calculated by improved Thornthwaite method [11]:

$$O_{ei}=\begin{cases}0 & T_i\le 1^{\circ}C\\ 1.6Day\left(\frac{10T_i}{I}\right)^{\alpha}\times 10 & 1<T_i\le 26.5^{\circ}C\\ \alpha_1+\alpha_2T_i+\alpha_3T_i^2 & T_i>26.5^{\circ}C\end{cases}.$$

In above formula, *Day* is the division of how many days of each month and 30, T_i the average temperature of the i^{th} month, α=0.49239+1.792 × 10^{-2}−7.71 × $10^{-5}I^2$+6.75 × $10^{-7}I^3$, $I=\sum_{1}^{12} i$ is the total heating index for 12 month, $i=\left(\frac{T_i}{5}\right)^{1.514}$ is the average heating index for each month. α_1= -415.8547, α_2=32.2441, α_3= -0.4325. According to calculation value, the range might be subdivided as {0.2-0.4, 0.4-0.6, 0.6-0.8, 0.8-1.0, 1.0-1.2} and the corresponding integer number range is {1,2,3,4,5}.

Finally, we obtained the value range of each factor (See Table 2).

TABLE 2 VALUE RANGE FOR EACH FACTOR

Label	Value Range				
	1	2	3	4	5
a_1	≤40%	40%-60%	60%-80%	80%-90%	≥90%
a_2	≤100	100-200	200-300	300-400	≥400
a_3	≤100	100-200	200-300	300-400	≥400
a_4	≥40%	30%-40%	20%-30%	10%-20%	≤10%
a_5	≥40%	30%-40%	20%-30%	10%-20%	≤10%
a_6	Very serious	Serious	Normal	Low	Inconsequential
a_7	≤100	100-200	200-300	300-400	≥400
a_8	≤10	10-15	15-20	20-30	≥30
a_9	≤0.4	0.4-0.6	0.6-0.8	0.8-0.9	≥0.9
a_{10}	Very low	Low	Normal	High	Very high
a_{11}	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0	1.0-1.2
a_{12}	≤40%	40%-60%	60%-80%	80%-90%	≥90%
d	Very poor	Poor	Normal	Good	Very good

III. ROUGH-SET APPROACH FOR COMPREHENSIVE AESSESSMENT OF URBAN WATER RESOURCES UTILIZATION

A. Necessary concepts of Rough set theory

Definition 1 Given decision information system IS as a quadruple ($U,C\cup D,V,f$). Let U be the Universe, the finite set of objects, $U=\{x_1, x_2,\ldots x_{|U|}\}$; C condition attributes set, $C\neq\Phi$, $C=\{a_1,a_2,\ldots,a_m\}$; D decision attributes set, $D\neq\Phi$, $C\cap D=\Phi$; Let V be the set of attributes values, such as, V_a the set of values of attributes a; $f{:}U\times C\to V$ is an information function which assigns particular values from domains of attributes to objects, such as $f(x,a)\in V_a$, for all $x\in U$ and $a\in C\cup D$.

Definition 2 Given IS = ($U,C\cup D,V,f$). To any equivalence relation C', i.e., a subset of condition attributes set C, $C'\subseteq C$, the Indiscernibility Relation $IND(S,C')$ could be defined as below:

$$IND(S,C')=\{(x,y)\mid (x,y)\in U^2, \forall\, c\in C', f(x,a)=f(y,a)\}$$

Obviously, $IND(S,C')$ is an equivalence relation. $IND(S,C')=\bigcup_{c\in C'} IND(S,\{c\})$. Let $U/IND(S, C')$ be the set of all equivalence class of $IND(S, C')$, marked as U/C'. Let $[x]_{C'}$ specify the equivalence class including the object x.

Hence, the equivalence classes including x decided by C and D could be expressed as $[x]_C$ and $[x]_D$ respectively. $[x]_C$ and $[x]_D$ contains all the equivalence class on C and D, i.e., $[x]_C=\{X_1,X_2,\ldots X_l\}$, $[x]_D=\{Y_1,Y_2,\ldots Y_{l'}\}$, $l,l'\le|U|$.

Definition 3 Given IS = ($U,C\cup D,V,f$). Let Y be the one of the equivalence classes of $[x]_D$, thus the classification on D can be expressed as $[x]^*_D=\{Y, U\text{-}Y\}=\{Y, \neg Y\}$. Then, $P(Y|[x]_C)$,the conditional probability of equivalence class Y on the equivalence class $[x]_C$ is:

$$P(Y\mid[x]_C)=\frac{|Y\cap[x]_C|}{|[x]_C|}\ [9].$$

Where $|[x]_C\cap Y|$ and $|[x]_C|$ represent the cardinality of each set. $[x]_C\subseteq Y$ means $[x]_C\subseteq[x]_D$, $[x]_C\cap Y\neq\Phi$ means $[x]_C$ might belongs to $[x]_D$. Thus, $P(Y|[x]_C)=1$ if and only if $[x]_C\subseteq Y$; $P(Y|[x]_C)>0$ if and only if $[x]_C\cap Y\neq\Phi$; $P(Y|[x]_C)=0$ if and only if $[x]_C\cap Y=\Phi$.

Definition 4 Given IS = ($U,C\cup D,V,f$) and $[x]_C$, $[x]_D$. Then the positive region of C on Y can be defined as:

$$POS(U,C,Y)=\bigcup_{P(Y|[x]_C)=1}\{x\in[x]_C\}.$$

The negative region is

$$NEG(U,C,Y)=\bigcup_{P(Y|[x]_C)=0}\{x\in[x]_C\}.$$

To any object $x\in U$, x belongs to class Y if and only if $x\in POS(U,C,Y)$; x belongs to class $\neg Y$ if and only if $x\in NEG(U,C,Y)$.

Definition 5 Given IS = ($U,C\cup D,V,f$). Let $c\in C$, if $POS(U,C,Y)=POS(U,C\text{-}\{c\},Y)$, then c is said to be *dispensable* in C with respect to class Y. If any of $c\in C$ is not dispensible, C is said to be *independent*. Otherwise, C is *dependent*. If a subset of condition attributes $C'\subseteq C$ is a minimal independent subset of condition attributes with repect to Y, it is said to be a *reduct*.Let RED(IS) be a set of all the reducts of IS.

Definition 6 Given IS = ($U,C\cup D,V,f$). The subset of all indispensable attribute(s) is the $CORE$ (IS). It also could be calculated by $CORE(IS)=\cap RED(IS)$.

CORE (IS) is the determinant set of original decision information system. It could be a base in calculating any reduct and it could be a null set.

Definition 7 Given IS = ($U,C\cup D,V,f$) and a reduct C'. The positive and negative decision rules from reduct C' are defined as below respectively:

$Des(x)\to Des(Y)$, if and only if $P(Y|[x]_C)=1$;

$Des(x)\to Des(\neg Y)$, if and only if $P(Y|[x]_C)=0$;

Where $Des(x)=\bigwedge_{a\in C'}(f(x,a))$; $Des(x)=f(x,d)$. $\wedge$ means the joint of each expression.

Definition 8 Given IS = ($U,C\cup D,V,f$), a reduct C' and a decision rule $Des(x)\to Des(Y)$, $x\in U$. If there is an undecided object x_1, the conditional probability of Y on the object x_1 with respect to x is:

$P_{x1}=|Des(x_1)|/|Des(x)|$.

Where $|Des(x_1)|$ and $|Des(x)|$ donate the cardinality of $Des(x_1)$ and $Des(x)$ respectively. Therefore, $P_{x1}=1$ if and only if $x_1 \subseteq Y$. If $P_{x1}=0$, it means $x_1 \subseteq \neg Y$.

B. Algorithm description

According to above definitions, the Rough Set Approach for the Evaluation of urban Water resources utilizaiton (RSAEW) was obtained.

RSAEW algorithm is described as below:

Input: Constructing an original decision information system IS with *m* condition attributes, 1 decision attribute and *n* objects according to Table 2 and Table 3 and collected data;

Output: Reduct *C'* with minimal numbers of conditional attributes, *CORE*(IS) and decision rules;

(a) $RED_j \leftarrow \Phi$; $j = 1$;

(b) for i =1 to m do

Random choose *C'*containing *i* attribute(s) from IS(i.e. C_m^i combinations,);

Calculate $POS(U,C,Y)$ and $POS(U,C',Y)$ according to Definition 4;

If $POS(U,C,Y)= POS(U,C',Y)$

$RED_j \leftarrow C'$; $j = j + 1$;

(c) Calculate *CORE*(IS) according to Definition 6;

(d) Comparing $| RED_j |$, Choose the minimum RED_j

Calculate all the decision rules of RED_j according to Definition 7;

(e) Using decision rules classify the undicided objects;

(f) Put the new object into IS;

Calculate step (b) and (c) again;

IV. CASE STUDY AND DISCUSSION

A. Study area

Beijing is a microcosm of China that may be one of the most seriously water-scarce metropolitan areas in the world. Its per capita annual average availability water is only 300 m^3, about 1/8 of the Chinese and 1/30 of the world average. Beijing lies in a typically water scarce region in Northern China situated in the center of the Haihe River Basin. It covers a total area of 1.64×10^4 km^2, and has five main rivers, named Yongding, Chaobai, Juma, Bei Canal, and Gou. Surface runoff and ground water are the main water sources. Beijing's total available water resources are low – the multi-year average is 3.8 billion m^3 and the mean annual precipitation is 585 mm An enormous population of over 18 million people therefore complicates provision and distribution of such scarce water supply. Unbalanced precipitation and considerable variance of total precipitation among years cause further complication [12].

B. Data acquisition

There are two way in collecting training samples, one is from local historical records, and another is from other similar areas. After analyzing Beijing's situation and past researches, we use local historical data of Beijing to construct training set. In this article, data include meteorological data (1986, 1988-2008, including temperature and precipitation data), social economical data (1986,1988-2008), and water resources statistic data (1986, 1988-2008) collected from China Meteorological Administration, Beijing Statistical Yearbook, Beijing Water Resources Bulletin and Beijing Environmental Statement respectively.

According to Table 2 and Beijing data, we obtained the training set (See Table 3) which will be treated as the original decision information system in RSAEW algorithm and the testing set (See Table 4).

TABLE 3 TRAINING SET SINCE 1986,1988-2005

Objects	Conditional Attributes Set												D
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
x_1	7	2	1	5	4	2	4	2	7	2	4	1	4
x_2	7	1	1	5	4	3	5	2	7	3	2	1	4
x_3	7	1	1	2	4	3	3	2	7	2	2	1	4
x_4	7	2	1	4	4	3	4	2	7	4	4	1	4
x_5	8	2	1	5	3	2	4	2	7	3	5	1	4
x_6	8	1	1	1	3	3	2	2	7	2	3	1	4
x_7	8	1	1	1	3	3	2	2	7	1	3	1	2
x_8	8	1	1	5	2	2	5	2	7	4	5	1	2
x_9	8	2	3	2	2	2	3	2	7	3	3	1	2
x_{10}	8	2	3	5	2	3	4	2	7	4	4	1	2
x_{11}	8	2	4	1	2	2	2	2	7	1	2	1	2
x_{12}	8	2	4	4	2	2	3	2	7	4	3	1	2
x_{13}	8	2	4	1	2	1	2	2	7	1	1	1	2
x_{14}	8	3	4	1	2	1	2	2	8	1	2	1	2
x_{15}	8	2	4	1	2	1	2	2	8	1	2	2	3
x_{16}	9	3	5	1	2	1	2	2	8	1	2	2	3
x_{17}	9	3	5	1	2	1	2	2	8	1	3	2	3
x_{18}	9	3	5	2	2	4	2	2	8	2	3	2	3
x_{19}	9	3	5	2	1	4	2	2	8	2	2	3	3

TABLE 4 TESTING SET 2006-2008

Objects	Conditional Attributes Set												D
	a_1	A_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
Obj_1	9	3	1	2	1	4	2	2	8	2	2	3	N/A
Obj_2	9	3	1	2	1	4	2	2	8	2	2	3	N/A
Obj_3	9	3	1	5	1	4	3	2	8	4	3	3	N/A

C. Data treatment and result analysis

According to RSAEW algorithm and Table 3, 20 reducts and the core of original decision information system, $\{a_{10}\}$, were calculated. And reduct *C'*, $\{a_{15}, a_{10}, a_{12}\}$, with the minimal attributes was presented (See Table 5). Obviously, the core set $\{a_{10}\}$, the determinant set of IS, indicates that the most important factor that affecting Beijing's water utilization status is annual precipitation of Beijing area which is seasoned with Beijing's location. Particularly, the annual precipitation has kept decreasing since 1999, while urban economy has been developing with remarkable annual increase. Large numbers of urban infrastructure construction and population explosion caused enormous water demand, sharp conflict between water supply and water demand, and large areas worsening pollution issues, especially under the current situation without any markedly progress in water use technology. These have also been represented by reduct *C'*.

TABLE 5 REDUCT *C'*

Objects	Conditional Attributes			D
	a_5	a_{10}	a_{12}	D
x_1,x_3	4	2	1	4
x_2	4	3	1	4
x_4	4	4	1	4
x_5	3	3	1	4
x_6	3	2	1	4
x_7	3	1	1	2
x_8,x_{10},x_{12}	2	4	1	2
x_9	2	3	1	2

x_{11},x_{13},x_{14}	2	1	1	2
x_{15},x_{16},x_{17}	2	1	2	3
x_{18}	2	2	2	3
x_{19}	1	2	3	3

From reduct *C'*, 12 positive decision rules (r_1 to r_{12}) presented as bellow:

r_1: $f(a_5,4)\wedge f(a_{10},2)\wedge f(a_{12},1) \Rightarrow f(d,4)$; r_2: $f(a_5,4)\wedge f(a_{10},3)\wedge f(a_{12},1) \Rightarrow f(d,4)$; r_3: $f(a_5,4)\wedge f(a_{10},4)\wedge f(a_{12},1) \Rightarrow f(d,4)$; r_4: $f(a_5,3)\wedge f(a_{10},3)\wedge f(a_{12},1) \Rightarrow f(d,4)$; r_5: $f(a_5,3)\wedge f(a_{10},2)\wedge f(a_{12},1) \Rightarrow f(d,4)$; r_6: $f(a_5,3)\wedge f(a_{10},1)\wedge f(a_{12},1) \Rightarrow f(d,2)$; r_7: $f(a_5,2)\wedge f(a_{10},4)\wedge f(a_{12},1) \Rightarrow f(d,2)$; r_8: $f(a_5,2)\wedge f(a_{10},3)\wedge f(a_{12},1) \Rightarrow f(d,2)$; r_9: $f(a_5,2)\wedge f(a_{10},1)\wedge f(a_{12},1) \Rightarrow f(d,2)$; r_{10}: $f(a_5,2)\wedge f(a_{10},1)\wedge f(a_{12},2) \Rightarrow f(d,3)$; r_{11}: $f(a_5,2)\wedge f(a_{10},2)\wedge f(a_{12},2) \Rightarrow f(d,3)$; r_{12}: $f(a_5,1)\wedge f(a_{10},4)\wedge f(a_{12},2) \Rightarrow f(d,3)$.

With respect to Table 4, the undecided objects, {Obj_1, Obj_2, Obj_3} could be decided by making use of r_{12}. Obviously, Obj_1 and Obj_2 will be classified into $f(d,3)$, which means that in 2006 and 2007, Beijing has kept the same status, {Normal}, with that of past five years. About Obj_3, it will be classified into $f(d,3)$ with the probability 2/3.

V. SUMAMARY AND CONCLUSION

With the deepening recognition of water resource utilization, evaluation method or model will be improved gradually. The evaluation algorithm RSAEW proposed in this article is based on Rough set theory and could discover rules only from urban historical data and hence assess urban water utilization status objectively. Meanwhile, the indicator system could be extended and the assessment also could be extended to a wide scopes.

Future research should give more attention to two aspects in particular:. (i) how to construct an all-round training set since the more comprehensive the training set is, the more authentic the result is; and (ii) area selection. For now, a city is taken as a research area because it is an area in which people live and for which it is relatively easy to obtain data. But water systems are clearly is not independent from their surrounding areas and are affected by geographical situation, local drainage area, and meteorological character. Therefore, how to choose an appropriate area with respect to all possible impact factors will be emphasized by researchers.

ACKNOWLEDGMENT

We are grateful to the Chinese Scholarship Council Joint Ph.D Project funded by Ministry of Education of P.R. China (Project No.2008637022) for its financial support of this research. Thanks to the helpful comments received from the anonymous reviewers and the editors.

REFERENCES

[1] Q.W. Ming, W.D. Yu, J.X. Zhang, “Fuzzy-based evaluation of water resources carrying capascity and its application,”Research of soil and water conservation, 2004,Vol.11,No.3, pp.14-17［闵庆文，余卫东，张建新. 区域水资源承载力的模糊综合评价分析方法与应用. 水土保持研究，2004,11 (3):14-17]

[2] X.M. Meng, H.P. Hu, “Application of set pair analysis model based on entropy weight to comprehensive evaluation of water quality,”Shuili Xuebao, 2009,Vol.40, No.3, pp.257-262 [孟宪萌, 胡和平. 基于熵权的集对分析模型在水质综合评价中的应用[J]. 水利学报, 2009, 40(3):257-262]

[3] X.Q.Wang, “A proposal and application of the intergrated benefit assessment model for urban water resources exploitation and utilization,”Water Resource Management, 2009,Vol.23,pp.1171-1182

[4] X.H. Yang, Z.F. Yang, Z. Y. Shen, “Multi-objective decision-making ideal interval method of the comprehensive evaluation of water renewability,”Science in China Ser.E Technological Sciences, 2004, Vol.34, Special Issue I, pp.34-41［杨晓华, 杨志峰, 沈珍瑶, 等. 水资源可再生能力综合评价的多目标决策理想区间法[J]. 中国科学 E 辑 技术科学，2004,34(增刊 I):34-41]

[5] Z. Y. Shen, Z. F.Yang, “Index system and method for assessing water resources renewability of the Yellow River Basin”. Journal of Natural Resources,2002,Vol.17, No.2, pp.188-197.[沈珍瑶，杨志峰. 黄河流域水资源可再生性评价指标体系与评价方法. 自然资源学报, 2002, 17(2):188-197.]

[6] T.Yang, Z.L.Wang, M.P.Xie, “Renewable evaluation of regional water resources based on rough set gan group eigenvalue method,” Journal of North A & F University (Nat. Sci. Ed.), 2008, Vol.36, No.1, pp.218-222［杨婷，王志良，谢敏萍等. 基于粗糙集和群决策特征根法的区域水资源可再生性评价. 西北农林科技大学学报（自然科学版），2008,36(1):218-222]

[7] X.Wang, D.S.Tang, “Evaluation method of regional water system based on rough set,” Programming and design of water conservancy, 2006, No.1, pp.31-34［王霞，唐德善. 基于粗糙模糊集的区域水资源系统的评价方法. 水利规划与设计，2006,1:31-34]

[8] An, A., Shan, N., Chan, C., Cercone, N. and Ziarko, W., “Discovering Rules for Water Demand Prediction: An Enhanced Rough-Set Approach”, Engineering Applications of Artificial Intelligence, Vol.9, No.6, 1996. pp.645-654

[9] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5):341-356.

[10] Y.Liu, A.Y. Hu, Y.Z.Deng, “Study on the evolution of water resource structure based on information entropy,” Journal of North A & F University (Nat. Sci. Ed.), 2006, Vol.34, No.6, pp.141-144[刘燕，胡安焱，邓亚芝. 基于信息熵的用水系统结构演化研究[J]. 西北农林科技大学学报（自然科学版），2006,34(6):141-144]

[11] J.Wei, Z.G. Ma, “Comparison of PalmerDrought Severity Index, percentage of precipitation anomaly and surface humid index,” Acta Geographica Sinica, 2003, Vol.58, Suppliement Issue, pp.117-124［卫捷, 马柱国. Palmer 干旱指数、地表湿润指数与降水距平的比较[J]. 地理学报，2003，58（增刊）：117-124.]

[12] Beijing Water Authority, “Capital Water Resources Sustainable Utilization Planning in the Beginning of 21st Century,” Beijing: China WaterPower Press, 2001