Pergamon

Engng Applic. Artif. Intell. Vol. 9, No. 6, pp. 645–653, 1996

PII: S0952-1976(96)00059-0 0952-1976/96 $15.00 + 0.00

Contributed Paper

Discovering Rules for Water Demand Prediction: An Enhanced Rough-set Approach

AIJUN AN
University of Regina, Canada

NING SHAN
University of Regina, Canada

CHRISTINE CHAN
University of Regina, Canada

NICK CERCONE
University of Regina, Canada

WOJCIECH ZIARKO
University of Regina, Canada

(Received April 1996)

Prediction of consumer demands is a pre-requisite for optimal control of water distribution systems because minimum-cost pumping schedules can be computed if water demands are accurately estimated. This paper presents an enhanced rough-sets method for generating prediction rules from a set of observed data. The proposed method extends upon the standard rough set model by making use of the statistical information inherent in the data to handle incomplete and ambiguous training samples. It also discusses some experimental results from using this method for discovering knowledge on water demand prediction.

Keywords: Water demand prediction, knowledge discovery, rough sets.

1. INTRODUCTION

The domain addressed is typical of a water distribution system of moderate-sized cities in North America. The sources of water are a lake and a number of underground wells. Water is pumped to reservoirs at a number of locations in the city, and is pumped from the reservoirs to the distribution system, or to another reservoir when it is necessary to adjust water levels. Pressures and rates of flow throughout the system can be controlled by means of pumps and valves housed in pumping stations. Human operators currently control operations of the distribution system at a central pumping station. The operators use heuristics or rules of thumb to minimize the cost of power used by pumps, to make demand forecasts and to keep the water level of reservoirs within reasonable ranges. These heuristics are based on a number of economic, environmental and sociological factors. Since the system is now controlled by a number of operators, it is difficult to standardize and optimize operations of the distribution system. Documenting the heuristics of the most experienced expert operator in an expert system is one way to reduce operating costs in the supply and distribution of purified water. In order to develop an expert system for the monitoring and control of the water distribution system, knowledge acquisition was conducted through structured and unstructured interviews with human experts, and heuristics were obtained for cost-effective water utility operations. Analysis of these

Correspondence should be sent to: Aijun An, Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2. E-mail: aijun@cs.uregina.ca

This paper has previously been published in *Proceedings of the International Joint Conference on Artificial Intelligence*. Copyright International Joint Conferences on Artificial Intelligences, Inc. Copies of this and other IJCAI Proceedings are available from Morgan Kaufmann Publishers, Inc., 340 Pine Street, 6th Floor, San Francisco, CA 94104, U.S.A. [http://www.mkp.com].

heuristics indicates that it is important that daily forecasts of water demand are accurately estimated, in order that minimum-cost pumping schedules may be derived. However, the prediction of water demand is currently poorly understood even by the experts, who approximate daily water demand based on their experience. The often inaccurate estimations result in inefficient operations of the water distribution system. The lack of knowledge on water demand prediction translates to a gap in the knowledge-base of the expert system. In other words, manual knowledge acquisition by itself is inadequate for handling all the situations that may arise in a complex engineering application.

An alternative method for knowledge acquisition is *automated discovery from observed data*, that is, to design an algorithm which can learn and refine decision rules from a set of training samples, or observed data. This method is referred to as *data mining* or *knowledge discovery*. This paper presents an application of a rough-set approach for the automated discovery of rules from a set of data samples for daily water demand predictions. The database contains training samples that cover information on environmental and sociological factors, and their corresponding daily volume of distribution flow. Since the training samples are incomplete and possibly ambiguous due to the characteristics of water demand and partial selection of condition factors which are described in Section 2, exact decision rules cannot be derived by standard methods.[1-4] The objective in this paper is to suggest a method for generating classification rules from incomplete information. The proposed method is based on an extension of the rough-set model.[5] Statistical information is used to define the positive and negative regions of a concept. Each classification rule generated by the learning system is characterized by a certainty factor which is in fact an estimate of the probability that an object matching the condition part of the rule belongs to the concept.

In the rest of the paper, the characteristics of water demand and knowledge representation are discussed in Section 2. Theoretical aspects of the method of rough set-based data analysis and rule generation are introduced in Sections 3 and 4. Experimental results from applying the method to the set of data samples are given in Section 5. The paper ends with a discussion of the method and future research issues.

2. DATA COLLECTION AND REPRESENTATION

2.1. Characteristics of water demand

The instantaneous consumption of water in an urban distribution system is determined by a large number of industrial, commercial, public and domestic consumers, distributed throughout the area supplied. This consumption is influenced by factors such as weather conditions, seasonal variation, day of the week and whether a particular day is a statutory holiday. Thus the total demand on an urban water distribution system is a time-varying, periodic, and nonstationary series, the modelling of which is difficult through computational methods alone.

Table 1. Condition factors for water demand prediction

Label	Condition attributes
a_1	Day of week
a_2	Today's maximum temperature
a_3	Today's minimum temperature
a_4	Today's average humidity
a_5	Today's rainfall
a_6	Today's snowfall
a_7	Today's average speed of wind
a_8	Yesterday's maximum temperature
a_9	Yesterday's minimum temperature
a_{10}	Yesterday's average humidity
a_{11}	Yesterday's rainfall
a_{12}	Yesterday's average speed of wind
a_{13}	Yesterday's bright sunshine hours
a_{14}	The day before yesterday's maximum temperature
a_{15}	The day before yesterday's average humidity
a_{16}	The day before yesterday's rainfall
a_{17}	The day before yesterday's average speed of wind
a_{18}	The day before yesterday's bright sunshine hours

2.2. Factor selection

Eighteen of the many factors that may affect the water consumption of a city have been selected because they are considered the more important ones (see Table 1). The first factor is the day of the week, which is chosen based on the observation that on weekends the daily total distribution flows are usually less than those on weekdays. The city also has the power, through by-laws, to restrict watering of lawns on Wednesdays. Furthermore, Mondays are known to be days of high water usage because many people do their laundry on Mondays, and in the summer people water the lawns on that day after returning from a weekend at their cottages. The other 17 factors are weather conditions on temperature, humidity, precipitation, wind and bright sunshine hours, grouped under three consecutive days. The values of these factors were obtained from a monthly meteorological summary from Environment Canada.

On the decision side, the historical information about water consumption has been recorded by the city, which calculates the daily water consumption by summing the daily distribution flows metered at each pumping station of the city. The summation reflects the total amount of water the city uses per day, which is needed by the operator at the central control station every day. The total value varies from 50 Ml on the coldest winter days to 180 Ml in summer.

2.3. Data representation

It is assumed that the given set of training samples represents the knowledge about the domain. In the approach described here, the training set is described by a *classification system*, also referred to as an *information system*.[3] The objects in a universe U are described by a set of attribute values.

Table 2. Classification system for water demand prediction

Objects	C														D
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	
obj_1	6	6	7	7	9	7	1	1	0	1	3	2	6	0	0
obj_2	1	5	7	6	9	9	0	0	0	0	1	5	1	0	0
obj_3	6	6	7	7	9	7	1	1	0	1	3	2	6	0	0
obj_4	3	7	7	7	5	2	0	0	0	0	1	3	1	2	1
obj_5	3	7	7	7	8	2	0	0	0	0	1	7	1	9	1
obj_6	3	7	7	7	5	2	0	0	0	0	1	3	1	2	1
obj_7	6	6	8	7	3	2	0	0	0	0	6	0	6	2	2
obj_8	3	7	7	7	5	2	0	0	0	0	1	3	1	2	2

Formally, a classification system S is a quadruple $\langle U, A, V, f\rangle$, where $U = \{x_1, x_2, \ldots, x_N\}$ is a finite set of objects, which in this case are states of the environment; A is a finite set of attributes; the attributes in A are further classified into two disjoint subsets, *condition* attributes C and *decision* attributes D, such that $A = C \cup D$ and $C \cap D = \emptyset$; $V = \bigcup_{a\in A} V_a$ is a set of attribute values and V_a is the *domain* of attribute a (the set of values of attribute a); $f: U \times A \to V$ is an information function which assigns particular values from domains of attributes to objects such that $f(x_i, a) \in V_a$, for all $x_i \in U$ and $a \in A$.

The classification system represents the classification of the states of the environment based on the values of attributes. The classification system for the set of observed samples for water demand prediction contains more than 300 objects, which consist of daily information on the condition factors and the decision attribute for 10 months from March to December 1994. In order to illustrate the theory used in the following sections, Table 2 shows eight of these objects projected on 14 condition attributes. For the purpose of rough-set-based data analysis, the classification system has been generalized by replacing the original attribute values with some discrete ranges, such as for example, attribute a_2 (minimum temperature) has been discretized into ten categories 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. The categories 7 and 8 for a_2 which appear in Table 2 stand for the ranges (6.46,10.34] and (10.34, 14.22] respectively*. It has to be emphasized at this point that the question of how to optimally discretize the attribute values is unsolved, and a subject of on-going research. Sections 3 and 4 will use Table 2 as an example of classification systems to illustrate the terminologies. The experimental results in Section 5 are based on the whole set of training samples.

3. DATA ANALYSIS

3.1. Indiscernibility relation

A classification system provides only partial information for characterizing subsets of the universe. That is, the set of selected attributes may not be sufficient to characterize the subsets of the universe unambiguously. Any two objects are indistinguishable from one another whenever they assume the same attribute values. This means that it may not be possible to distinguish all the objects solely by means of the admitted attributes and their values.

Given a classification system $\langle U, A, V, f\rangle$, let B be a subset of A, and let x_i and x_j be members of U. A binary relation $R(B)$, called an *indiscernibility relation*, is defined as $R(B) = \{(x_i, x_j) \in U^2 \mid \forall a \in B,\ f(x_i, a) = f(x_j, a)\}$. x_i and x_j are said to be indiscernible by the set of attributes B in S if $f(x_i, a) = f(x_j, a)$ for every $a \in B$. For example, in the classification system shown in Table 2, obj_1 and obj_3 are indiscernible by the set of attributes $C \cup D$; obj_4, obj_6 and obj_8 are indiscernible by the set of condition attributes C.

Clearly, $R(B)$ is an equivalence relation on U for every $B \subseteq A$. Thus, two natural equivalence relations $R(C)$ and $R(D)$ can be defined on U for an information system S. A *concept* Y is an equivalence class of the relation $R(D)$. The objective is to construct decision rules for each concept. Given a concept Y, the partition of U with respect to this concept is defined as $R^*(D) = \{Y, U - Y\} = \{Y, \neg Y\}$.

Based on the set of condition attributes C, an object x_i specifies the equivalence class $[x_i]_R$ of the relation $R(C)$:

$$[x_i]_R = \{x_j \in U \mid \forall a \in C,\ f(x_j, a) = f(x_i, a)\}.$$

It can be said that x_i *definitely* belongs to a concept Y if $[x_i]_R \subseteq Y$ and that x_i *possibly* belongs to the concept Y if $[x_i]_R \cap Y \neq \emptyset$.

The conditional probability of a concept Y on the equivalence class $[x_i]_R$ can be estimated as

$$P(Y \mid [x_i]_R) = \frac{|Y \cap [x_i]_R|}{|[x_i]_R|},$$

where $|Y \cap [x_i]_R|$ and $|[x_i]_R|$ denote the number of objects in $Y \cap [x_i]_R$ and $[x_i]_R$ respectively. Therefore, $P(Y \mid [x_i]_R) = 1$ if and only if $[x_i]_R \subseteq Y$; $P(Y \mid [x_i]_R) > 0$ if and only if $[x_i]_R \cap Y \neq \emptyset$; and $P(Y \mid [x_i]_R) = 0$ if and only if $[x_i]_R \cap Y = \emptyset$.

Example 3.1

Consider the classification system given in Table 2. The concepts in this classification system, i.e. the equivalence classes on the relation $R(D)$, are

*The unmatched brackets stand for a half open and half closed range. For example, 6.46 does not belong to the range (6.46,10.34], but 10.34 does.

$$Y_0 = \{obj_1, obj_2, obj_3\}$$
$$Y_1 = \{obj_4, obj_5, obj_6\}$$
$$Y_2 = \{obj_7, obj_8\}.$$

The equivalence classes on the relation $R(C)$ are

$$X_1 = [obj_1]_R = [obj_3]_R = \{obj_1, obj_3\}$$
$$X_2 = [obj_2]_R = \{obj_2\}$$
$$X_3 = [obj_4]_R = [obj_6]_R = [obj_8]_R = \{obj_4, obj_6, obj_8\}$$
$$X_4 = [obj_5]_R = \{obj_5\}$$
$$X_5 = [obj_7]_R = \{obj_7\}.$$

Since $X_4 \subseteq Y_1$, obj_5 definitely belongs to the concept Y_1. The objects obj_4, obj_6, and obj_8 possibly belong to the concept Y_1 because the intersection of their equivalence class X_3 and Y_1 is not empty. Other objects do not belong to Y_1.

The conditional probability of the concept Y_1 on each equivalence class of $R(C)$ is as follows:

$$P(Y_1|X_1) = 0$$
$$P(Y_1|X_2) = 0$$
$$P(Y_1|X_3) = \frac{2}{3} = 0.667$$
$$P(Y_1|X_4) = 1$$
$$P(Y_1|X_5) = 0.$$

3.2. β-probabilistic approximation classification

Given a classification system $S = \{U, A, V, f\}$ and an equivalence relation $R(C)$ (*an indiscernibility relation*) on U, an ordered pair $AS = \langle U, R(C)\rangle$ is called an *approximation space*[3] based on the condition attributes C. The equivalence classes of the relation $R(C)$ are called *elementary sets* in AS because they represent the smallest groups of objects which are distinguishable in terms of the attributes and their values. Let $Y \subseteq U$ be a subset of objects representing a concept, and $R^*(C) = \{X_1, X_2, \ldots, X_n\} = \{[x_1]_R, [x_2]_R, \ldots, [x_n]_R\}$ be the collection of equivalence classes induced by the relation $R(C)$. In the standard rough set model, the *lower* and *upper* approximations of a set Y are defined by:

$$\underline{R(C)}(Y) = \bigcup_{P(Y|X_i)=1} \{X_i \in R^*(C)\} \quad \text{and}$$

$$\overline{R(C)}(Y) = \bigcup_{P(Y|X_i)>0} \{X_i \in R^*(C)\},$$

respectively. These definitions do not make use of the statistical information in the boundary region $\overline{R(C)}(Y) - \underline{R(C)}(Y)$. For this reason, an attempt is made to rectify this limitation by introducing a β-approximation space, which is essentially described in Refs 5 and 6.

A *β-approximation space* AS_P is a triple $\langle U, R(C), P\rangle$, where P is a probability measure described in Section 3.1 and β is a real number in the range $(0.5, 1]$. The β-approximation of a set Y in the space AS_P can be expressed in the following regions:

(1) β-positive region of the set Y: $POS_C^{\beta}(Y) = \bigcup_{P(Y|X_i)\geq\beta}\{X_i \in R^*(C)\}$.
(2) β-boundary region of the set Y: $BND_C^{\beta}(Y) = \bigcup_{1-\beta<P(Y|X_i)<\beta}\{X_i \in R^*(C)\}$.
(3) β-negative region of the set Y: $NEG_C^{\beta}(Y) = \bigcup_{P(Y|X_i)\leq 1-\beta}\{X_i \in R^*(C)\}$.

Clearly, the β-positive region of the set Y corresponds to all those elementary sets of U which can be classified into the concept Y with conditional probability $P(Y|X_i)$ greater than or equal to the parameter β. Similarly, the negative region of the set Y corresponds to all those elementary sets of U which can be classified into the set $\neg Y$ with the probability $P(\neg Y|X_i) \geq \beta$.

Let $x_i \in U$ be an object; $POS_C^{\beta}(Y)$ and $NEG_C^{\beta}(Y)$ are the positive and negative regions of the concept Y respectively. The object x_i is classified belonging to the concept Y if and only if $x_i \in POS_C^{\beta}(Y)$ or classified belonging to the complement $\neg Y$ of the concept Y if and only if $x_i \in NEG_C^{\beta}(Y)$. In fact, it is necessary to decide whether x_i is in the concept Y on the basis of the set of equivalence classes in AS_P rather than on the basis of the set Y. This means dealing with $POS_C^{\beta}(Y)$ and $NEG_C^{\beta}(Y)$ instead of the set Y. One can see that if x_i is in $POS_C^{\beta}(Y)$, it can be classified into the concept Y with conditional probability $P(Y|X_i)$ greater than or equal to the parameter β. If x_i is in $NEG_C^{\beta}(Y)$, then it can be classified into $\neg Y$ with the probability $P(\neg Y|X_i)$ greater than or equal to β. If the object is in the boundary region, then the classification can be made either into Y or $\neg Y$ with the probability less than β.

Example 3.2

Following from Example 3.1, recall $Y_1 = \{obj_4, obj_5, obj_6\}$.

(1) Let $\beta = 1$. β-positive region of the set Y_1:
$$POS_C^{\beta}(Y_1) = \{obj_5\};$$
β-negative region of the set Y_1:
$$NEG_C^{\beta}(Y_1) = \{obj_1, obj_2, obj_3, obj_7\}.$$
β-boundary region of the set Y_1:
$$BND_C^{\beta}(Y_1) = \{obj_4, obj_6, obj_8\}.$$
(2) Let $\beta = 0.6$. β-positive region of the set Y_1:
$$POS_C^{\beta}(Y_1) = \{obj_4, obj_5, obj_6, obj_8\};$$
β-negative region of the set Y_1:
$$NEG_C^{\beta}(Y_1) = \{obj_1, obj_2, obj_3, obj_7\}.$$
β-boundary region of the set Y_1:
$$BND_C^{\beta}(Y_1) = \varnothing.$$

3.3. Reduction of condition attributes

In a classification system there often exist some condition attributes that do not provide any additional information about the objects in U. It is desirable to remove those attributes, since the complexity and cost of a decision process can be reduced if those condition attributes are eliminated. In this subsection, the concept of *reduct* in rough sets is used to describe the method of condition attribute reduction.

Given an attribute-value system $S = \langle U, C \cup D, V, f \rangle$, an attribute a is said to be *dispensable* in C with respect to a concept Y if $POS^{\beta}_{C-\{a\}}(Y) = POS^{\beta}_{C}(Y)$; otherwise a is an *indispensable* attribute in C with respect to Y. A subset of condition attributes $B \subseteq C$ is said to be a *dependent set* in S with respect to Y if there exists a proper subset $K \subset B$ such that $POS^{\beta}_{B}(Y) = POS^{\beta}_{K}(Y)$; otherwise B is an *independent set* with respect to Y. A *reduct* C' of attributes C is a maximal independent subset of condition attributes with respect to Y.

The procedure for finding a single reduct is very straightforward. Consider a condition attribute $a \in C$. If the β-positive region $POS^{\beta}_{C-\{a\}}(Y)$ of the set Y is the same as $POS^{\beta}_{C}(Y)$, then the attribute a is marked as being redundant and is removed from the set of condition attributes C. Other superfluous condition attributes can be removed in the same manner. The remaining set of condition attributes is a reduct. More than one reduct may exist for a given attribute-value system. Selection of a "best" reduct depends on the optimality criterion associated with the attributes. Significance values can be assigned to attributes, and the selection is based on those values.

Example 3.3

The condition attributes C of the information system in Table 2 has a total of 20 reducts. A reduced information system based on the reduct $\{a_2, a_3, a_7\}$ with respect to the concept Y_1 is shown in Table 3. The objects with value '1' for column Y_1 belong to β-positive region of the concept Y_1; the objects with value '0' for column Y_1 belong to β-negative region of concept Y_1. Here $\beta = 0.6$.

4. RULE GENERATION

Rule generation is a crucial task in any learning system. This section describes how decision rules are generated, based on the reduct obtained from Section 3.

4.1. Generating probabilistic decision rules

Let $R^*(RED) = \{X_1, X_2, \ldots, X_n\}$ be the collection of equivalence classes of the relation $R(RED)$ where RED is a reduct which is a reduced set of condition attributes C in S, and let $R^*(D) = \{Y, \neg Y\}$ be the partition induced by the decision attribute. Each equivalence class X_i of the equivalence relation $R(RED)$ is associated with a unique combination of values of attributes belonging to RED. This combination of values is referred to as the *description* of the equivalence class $X_i \in R^*(RED)$. The description of X_i can be expressed as:

$$Des(X_i) = \bigwedge_{a \in RED} (a = f(x_i, a)),$$

where $\bigwedge$ denotes the conjunction operator, and x_i is an object in the equivalence class X_i. Similarly, the descriptions of Y and $\neg Y$ are:

$$Des(Y) = (d = f(x_i, d))$$

and

$$Des(\neg Y) = (d \neq f(x_i, d)),$$

where d is the decision attribute in D and $x_i \in Y$. Without loss of generality, D is considered as a singleton set.

The relationship between the partition $R^*(RED)$ and the partition $R^*(D)$ can be described by the following decision rules: for $X_i \in R^*(RED)$,

(1) $Des(X_i) \rightarrow^{c_i} Des(Y)$, if $P(Y|X_i) \geq \beta$,
(2) $Des(X_i) \rightarrow^{c_i} Des(\neg Y)$, if $P(Y|X_i) \leq 1 - \beta$,

where c_i is a certainty factor, which is equal to $P(Y|X_i)$ in the first case and $1 - P(Y|X_i)$ in the second. This means that if an object x_i satisfies the description $Des(X_i)$ and if $P(Y|X_i) \geq \beta$, then the object x_i "belongs" to Y with certainty c_i. Similarly, if $P(Y|X_i) \leq 1 - \beta$, then the object x_i "belongs" to the complementary concept $\neg Y$ with certainty c_i.

Example 4.1

Following from Example 3.3, let RED denote the reduct $\{a_2, a_3, a_7\}$. The collection $R^*(RED)$ of the equivalence classes of the relation $R(RED)$ is

$$R^*(RED) = \{X_1, X_2, X_3, X_4\} = \{\{obj_4, obj_5, obj_6, obj_8\}, \{obj_1, obj_3\}, \{obj_2\}, \{obj_7\}\}.$$

The descriptions of these equivalence classes are as follows:

$$Des(X_1) = (a_2 = 7) \wedge (a_3 = 7) \wedge (a_7 = 0)$$
$$Des(X_2) = (a_2 = 7) \wedge (a_3 = 7) \wedge (a_7 = 1)$$
$$Des(X_3) = (a_2 = 7) \wedge (a_3 = 6) \wedge (a_7 = 0)$$
$$Des(X_4) = (a_2 = 8) \wedge (a_3 = 7) \wedge (a_7 = 0).$$

The descriptions of Y_1 and $\neg Y_1$ are

$$Des(Y_1) = (d = 1)$$
$$Des(\neg Y_1) = (d \neq 1).$$

Since $Y_1 = \{obj_4, obj_5, obj_6\}$, the condition probabilities are calculated as

Table 3. Reduct table with respect to Y_1 after reduction from Table 2

Objects	C			Y_1	Number of objects in Y_1	Number of objects in $\neg Y_1$
	a_2	a_3	a_7			
$obj_4, obj_5, obj_6, obj_8$	7	7	0	1	3	1
obj_1, obj_3	7	7	1	0	0	2
obj_2	7	6	0	0	0	1
obj_7	8	7	0	0	0	1

$$P(Y_1|X_1) = \frac{3}{4} = 0.75$$
$$P(Y_1|X_2) = 0$$
$$P(Y_1|X_3) = 0$$
$$P(Y_1|X_4) = 0.$$

The decision rules with respect to the concept Y_1 are then obtained as follows:

$$r_1^+ : (a_2 = 7) \wedge (a_3 = 7) \wedge (a_7 = 0) \rightarrow^{0.75} (d = 1)$$
$$r_1^- : (a_2 = 7) \wedge (a_3 = 7) \wedge (a_7 = 1) \rightarrow^{1} (d \neq 1)$$
$$r_2^- : (a_2 = 7) \wedge (a_3 = 6) \wedge (a_7 = 0) \rightarrow^{1} (d \neq 1)$$
$$r_3^- : (a_2 = 8) \wedge (a_3 = 7) \wedge (a_7 = 0) \rightarrow^{1} (d \neq 1).$$

4.2. Rule generalization

As indicated in the previous subsection, decision rules can be directly obtained from the *reduced* information system. One rule is obtained for each equivalence class of the partition $R^*(RED)$. However, they may contain attributes whose values are irrelevant for determining the target concept. Furthermore, the rules can be further generalized by inspecting which conditions in a rule can be removed without causing any inconsistency. A decision rule obtained by dropping the maximum possible number of conditions is called a "maximally general" rule. By construction, the maximally general rules contain a minimum number of conditions. The *decision matrix* technique[7] is used here to find all the maximally general rules.

Let X_i^+, $i = (1, 2, \ldots, \gamma)$, denote the equivalence classes of the relation $R^*(RED)$ such that $X_i^+ \subseteq POS_{RED}^{\beta}(Y)$, and let X_j^-, $j = (1, 2, \ldots, \rho)$, denote the equivalence classes of the relation $R^*(RED)$ such that $X_j^- \subseteq NEG_{RED}^{\beta}(Y)$. A decision matrix $M = (M_{ij})_{\gamma\times\rho}$ is defined by:

$$M_{ij} = \{(a, f(X_i^+, a)) : a \in RED, f(X_i^+, a) \neq f(X_j^-, a)\}$$

where a is a condition attribute belonging to RED. That is, the entry M_{ij} contains all attribute-value pairs whose values are not the same between the equivalence class X_i^+ and the equivalence class X_j^-. The set of decision rules computed for a given equivalence class X_i^+ is obtained by treating each element of M_{ij} as a Boolean expression and constructing the following Boolean function, namely:

$$B_i = \bigwedge_j (\bigvee M_{ij}),$$

Table 4. The decision matrix for the rules with respect to concept Y_1

	X_1^-	X_2^-	X_3^-
X_1^+	$(a_7, 0)$	$(a_3, 7)$	$(a_2, S7)$

where $\bigwedge$ and $\bigvee$ are the usual conjunction and disjunction operators.

It can be shown that the prime implicants of the Boolean function B_i are in fact the maximally general rules for the equivalence class X_i^+ belonging to the positive learning region $POS_{RED}^{\beta}(Y)$. Thus, by finding the prime implicants of all the decision functions B_i ($i = 1, 2, \ldots, \gamma$), all the maximally general rules can be computed for the positive learning region $POS_{RED}^{\beta}(Y)$.

Example 4.2

Following from Example 4.1, the equivalence classes of the relation $R(RED)$ are listed as follows:

$$X_1^+ = X_1 = \{obj_4, obj_5, obj_6, obj_8\}$$
$$X_1^- = X_2 = \{obj_1, obj_3\}$$
$$X_2^- = X_3 = \{obj_2\}$$
$$X_3^- = X_4 = \{obj_7\}.$$

The decision matrix $M = (M_{ij})_{1\times 3}$ is shown in Table 4. The Boolean expression B_1 is then obtained as $(a_7 = 0) \wedge (a_3 = 7) \wedge (a_2 = 7)$. Since the prime implicants of B_1 is B_1 itself, i.e. $(a_7 = 0) \wedge (a_3 = 7) \wedge (a_2 = 7)$, the maximally general rule for the equivalence class X_1^+ is

$$(a_7 = 0) \wedge (a_3 = 7) \wedge (a_2 = 7) \rightarrow^{0.75} (d = 1)$$

which is identical to the rule r_1^+ in Example 4.1.

Given the set of all maximally general rules for an information system S, the system provides the options of finding the set of *minimal rules* and the set of *minimal covering rules*. The support set of a rule r_i, denoted as $supp(r_i)$, is defined as the collection of rows of the original table satisfying the condition part of r_i. A collection of rules RUL' is said to be a set of *minimal rules*, if for every $r_i \in RUL'$, $supp(r_i) \nsubseteq supp(r_j)_{\forall r_j \in RUL', r_i \neq r_j}$. A collection of rules RUL'' is said to be a set of *minimal covering rules*, if for every $r_i \in RUL''$, $supp(r_i) \nsubseteq \bigcup supp(r_j)_{r_j \in RUL'', r_i \neq r_j}$.

Example 4.3

As shown in Example 4.2, the maximally general rule for the concept Y_1 can be calculated through a decision matrix. Similarly, the maximally general rules for concepts Y_0 and Y_2 can be obtained. The set of all maximally general rules for the information system in Table 2 are

$r_1 : (a_7 = 1) \rightarrow^1 (d = 0)$

$r_2 : (a_3 = 6) \rightarrow^1 (d = 0)$

$r_3 : (a_2 = 7) \wedge (a_3 = 7) \wedge (a_7 = 0) \rightarrow^{0.75} (d = 1)$

$r_4 : (a_2 = 8) \rightarrow^1 (d = 2)$

where r_1 and r_2 are for the concept Y_0, r_3 is for the concept Y_1, and r_4 is for the concept Y_2.

The support sets of these rules are

$$supp(r_1) = \{obj_1, obj_3\}$$
$$supp(r_2) = \{obj_2\}$$
$$supp(r_3) = \{obj_4, obj_5, obj_6, obj_8\}$$
$$supp(r_4) = \{obj_7\}.$$

Since these support sets are not overlapping, the rules are both minimal rules and minimal covering rules.

5. EXPERIMENTAL RESULTS

The method presented here has been implemented in the GRG system[8] developed at the University of Regina. The system addresses the multiple tasks of data analysis, database mining, pattern recognition and validation, and expert-system building. The proposed method was tested by applying the GRG system on a set of recorded data for water demand prediction. The objective is to analyze a set of training data and generate a set of decision rules. Such decision rules can be used to predict a city's daily water demand. As discussed in Section 2, the set of training samples consists of more than 300 objects collected over a period of ten months, and includes information on day of the week, weather conditions and daily consumption of water.

In the experiment, the values of the decision attribute were divided into ten ranges, so that the information system has ten concepts. Samples of the generated rules are shown below.

The most general rule for the concept $D = [53 - 63]$ is

$$(-6.70 < a_3 \leq -1.74) \wedge (< -6.70 < a_9 \leq -1.74) \wedge (11.86 < a_14 \leq 17.42) \rightarrow^1 (53 < D \leq 63).$$

This rule covers 25% of the training objects concluding the concept. The rule states that: if today's and yesterday's minimum temperatures are both between −6.70 and −1.74°*C* and the day before yesterday's maximum temperature is between 11.86 and 17.42°*C*, then the water demand is between 53 and 63 Ml with a certainty factor being 1.

The most general rule for the concept $D = (63 - 73]$ is

$$(81 < a_{10} \leq 87) \rightarrow^1 (63 < D \leq 73).$$

This rule covers 12% of the training objects concluding the concept. The rule states that: if yesterday's average humidity is between 81 and 87%, then the water demand is between 63 and 73 Ml with a certainty factor being 1.

The most general rule for the concept $D = (73 - 84]$ is

$$(8.7 < a_7 \leq 12.9) \wedge (-1.74 < a_9 \leq 3.22) \rightarrow^1 (73 < D \leq 84).$$

This rule covers 6.25% of the training objects concluding the concept. It states that: if today's average speed of wind is between 8.7 and 12.9 km/h and yesterday's minimum temperature is between −1.74 and 3.22°*C*, then the water demand is between 73 and 84 Ml with a certainty factor being 1.

A rule for the concept $D = (94 - 104]$ is

$$(47 < a_4 \leq 53) \wedge (17.42 < a_8 \leq 22.98) \wedge (3.22 < a_9 \leq 8.18) \rightarrow^1 (94 < D \leq 104).$$

This rule covers 17.6% of the training objects concluding the concept. It states that: if today's average humidity is between 47 and 53% and yesterday's maximum temperature is between 17.42 and 22.98°*C* and yesterday's minimum temperature is between 3.22 and 8.18°*C*, then the water demand is between 94 and 104 Ml with a certainty factor being 1.

The most general rule for the concept $D = (104 - 114]$ is

$$(22.98 < a_2 \leq 28.54) \wedge (53 < a_{15} \leq 58) \wedge (4.5 < a_{17} \leq 8.7) \rightarrow^1 (104 < D \leq 114).$$

This rule covers 33.3% of the training objects concluding the concept. It states that: if today's maximum temperature is between 22.98 and 28.54°*C* and the day before yesterday's average humidity is between 53 and 58% and the day before yesterday's average speed of wind is between 4.5 and 8.7 km/h, then the water demand is between 104 and 114 Ml with a certainty factor being 1.

The most general rule for the concept $D = (114 - 124]$ is

$$(36 < a_{10} \leq 41) \wedge (36 < a_{15} \leq 41) \rightarrow^1 (114 < D \leq 124).$$

This rule covers 33.3% of the training objects concluding the concept. It states that: if yesterday's and the day before yesterday's average humidities are both between 36 and 41%, then the water demand is between 114 and 124 Ml with a certainty factor being 1.

The most general rule for the concept $D = (124 - 134]$ is

$$(53 < a_4 \leq 58) \wedge (22.98 < a_{14} \leq 28.54) \wedge (13.30 < a_{18} \leq 15.20) \rightarrow^1 (124 < D \leq 134).$$

This rule covers 66.7% of the training objects concluding the concept. It states that: if today's average humidity is between 53 and 58% and the day before yesterday's maximum temperature is between 22.98 and 28.54°*C* and the day before yesterday's bright sunshine hours are between 13.30 and 15.20 hours, then the water demand is between 124 and 134 Ml with a certainty factor being 1.

In order to evaluate the rules derived by the new method, a *Leave-Ten-Out* experiment was conducted by using 90% of the data for training and the remaining 10% of data for testing. The error rate depends on the selection of training samples. The experiment was conducted ten times. The best error rate of prediction is 6.67% and the average error rate of prediction is 10.27%.

6. CONCLUSION

This paper has suggested a method for generating prediction rules from a given set of training examples. The proposed method extends upon the standard rough-sets method, and makes use of the statistical information inherent in the knowledge system. In this way, the method is capable of deriving imprecise decision rules with decision probabilities. This capability is important in situations when complete and deterministic information in empirical data is unavailable.

Application of this knowledge discovery method for water demand prediction has the advantage that important relationships between condition factors and the decision variables of water consumption are described in terms of simple "if–then" rules, which can be easily understood by users. This method is general and can also be applied to other domains, such as general consumer demand prediction, fault diagnosis and process control. The results reported in Section 5 are based on an earlier recorded data set. Items on a future research agenda include collecting more data to improve the accuracy of the prediction rules, validation of the rules by human experts, and incorporating the rules into the knowledge base for monitoring and control of the water distribution system.

Acknowledgements—This research is supported by Telecommunications Research Laboratories (TRLabs) and the Natural Science and Engineering Research Council of Canada. The authors would also like to thank Dr Paitoon Tontiwachwuthikul for his invaluable suggestions on this work.

REFERENCES

1. Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (eds) *Machine Learning: An Artificial Intelligence Approach*, Vol. 1. Morgan Kaufmann, San Mateo, CA, 1983.
2. Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds) *Machine Learning: An Artificial Intelligence Approach*, Vol. 2, Morgan Kaufmann, San Mateo, CA (1986).
3. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers (1991).
4. Quinlan, J. R., Induction of decision trees. *Machine Learning*, 1986, **1**, 81–106.
5. Ziarko, W., Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, 1993, **46**(1), 39–59.
6. Katzberg, J. and Ziarko, W., Variable precision rough sets with asymmetric bounds. In *Rough Sets, Fuzzy Sets and Knowledge Discovery*, ed. W. Ziarko. Springer, Berlin, 1994, 167–177.
7. Ziarko, W. and Shan, N., A rough set-based method for computing all minimal deterministic rules in attribute-value systems. University of Regina Tech. Report CS-93-02, 1993.
8. Shan, N., Hamilton, H. J. and Cercone, N., GRG: knowledge discovery using information generalization, information reduction, and rule generation. *Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence (ICTAI'95)*, Washington, DC, 1995, 372–379.

AUTHORS' BIOGRAPHIES

Aijun An is a Ph.D. candidate in Computer Science at the University of Regina. Her research interests include machine learning, case-based reasoning, knowledge engineering for developing expert systems, validation and verification of knowledge-based systems and applications of AI techniques to practical problems. She received a B.Sc. in computer science in 1987 and an M.Sc. in computer science in 1990 both from Xidian University, Xi'an, China. She joined the University of Regina in September 1993. She has published 16 papers in refereed journals and conference proceedings. She is a student member of IEEE, AAAI and the IEEE Computer Society.

Ning Shan is a Ph.D. candidate in Computer Science at the University of Regina. His research interests include machine learning, rough sets, rule-based expert systems, knowledge discovery in databases and real-world applications of artificial intelligence. He received a B.Eng. in Computer Engineering in 1983 from ChengDu Electronic Engineering Institute, China. He joined the University of Regina in 1993 and received an M.Sc. in Computer Science in 1995. He has published more than 30 papers in refereed books, journals and conference proceedings. He has developed the KDD-R, GRG and DBROUGH knowledge discovery systems at the University of Regina.

Christine W. Chan has been an assistant professor in the Department of Computer Science of the University of Regina since January 1993. Her current research interests include knowledge engineering for intelligent system construction, knowledge acquisition, technology-assisted learning, and industrial applications of artificial intelligence. She graduated with a Ph.D. in the interdisciplinary studies of Computer Science, Philosophy, and Psychology from Simon Fraser University in 1992. Dr Chan also has Master's degrees in Computer Science and Management Information Systems from the University of British Columbia and in History from the University of Wisconsin-Madison. Her bachelor's degree was obtained from Stanford University. She has published more than 40 refereed journal or conference papers, and is a member of AAAI, the IEEE Computer Society, and the Canadian Information Processing Society (CIPS).

Nick Cercone is the Associate Vice President (Research), Dean of Graduate Studies and the International Liaison Officer at the University of Regina. Formerly he was Director of the Centre for Systems Science at Simon Fraser University, and past chairman of the School of Computing Science. He received a B.S. degree in Engineering Science from the University of Steubenville in 1968, an MS degree in Computer and Information Science from Ohio State University in 1970, and a Ph.D. degree in Computing Science from the University of Alberta in 1975. He worked for the IBM Corporation in 1969 and 1971 on engineering design automation. His research interests include design and human interfaces, knowledge-based systems, knowledge-discovery in databases, and natural language processing. He has authored over 140 refereed technical papers, and is founder and co-editor of *Computational Intelligence*, is a member of the ACM, IEEE, AAAI, AISB, AGS and ACL, and a past president of the CSCSI/SCEIO (Canadian Society for Computational Studies of Intelligence).

Wojciech Ziarko received an M.Sc. degree in Applied Mathematics, specializing in the theory of information systems from the Warsaw University of Technology, Warsaw, Poland, in 1975. In 1980, he received a Ph.D. degree in Computer Science, specializing in databases, from the Institute of Computer Science of the Polish Academy of Sciences, Warsaw, Poland. In 1982, he joined the University of Regina, Canada where he is now Professor in the Computer Science Department. His research interests are data mining, machine learning, pattern classification and control-algorithm acquisition from data. These research interests are to a large degree motivated by the recent introduction of the theory of rough sets, which serves as a basic philosophical and theoretical framework in much of his research work. He has authored or co-authored almost a hundred research papers, edited one book on the above subjects, and organized the International Workshop on Rough Sets and Knowledge Discovery (Banff, 1993). Dr Ziarko designed several systems for rough set-based data analysis including systems DataQuest, DataLogic and KDD-R. Currently, he is heavily involved in the development of applications of his research in areas such as market data analysis and environmental control.