Applying Knowledge Discovery to Predict Water-Supply Consumption

**Aijun An, Christine Chan, Ning Shan, Nick Cercone, and Wojciech Ziarko,
University of Regina, Saskatchewan, Canada**

OPTIMIZING CONTROL OF OPERations in a municipal water-distribution system can reduce electricity costs and realize other economic benefits. However, optimal control requires an ability to precisely predict short-term water demand so that minimum-cost pumping schedules can be prepared. One of the objectives of our project to develop an intelligent system for monitoring and controlling municipal water-supply systems is to ensure optimal control and reduce energy costs. Hence, prediction of water demand is essential. In this article, we present an application of a rough-set approach for automated discovery of rules from a set of data samples for daily water-demand predictions. The database contains 306 training samples, covering information on seven environmental and sociological factors and their corresponding daily volume of distribution flow.

A ROUGH-SET METHOD GENERATES PREDICTION RULES FROM OBSERVED DATA, USING STATISTICAL INFORMATION INHERENT IN THE DATA TO HANDLE INCOMPLETE AND AMBIGUOUS TRAINING SAMPLES. EXPERIMENTAL RESULTS INDICATE THAT THIS METHOD PROVIDES MORE PRECISE INFORMATION THAN IS AVAILABLE THROUGH KNOWLEDGE ACQUISITION FROM HUMAN EXPERTS.

The problem domain

The problem domain is typical of a water-distribution system of moderate-sized cities in North America. The sources of water are a lake and several underground wells. Water is first pumped to reservoirs at various locations in the city and then from the reservoirs to the distribution system, or to another reservoir when it is necessary to adjust water levels.

Pumps and valves, housed in pumping stations, control the system's pressures and flow rates. Human operators control the distribution system's operations 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 depend on several economic, environmental, and sociological factors. A sample heuristic is,

> If the weather in the last three days is hot and dry, and the weather in the next three days is expected to be hot and dry, and the time before high demand is expected to be less than or equal to eight hours, then use a large pump and run it for a short time.

Because several operators control the system, standardizing and optimizing the distribution system's operations is difficult. Documenting the heuristics of the most experienced operator in an intelligent system is one way to reduce operating costs in the supply and distribution of purified water.

To develop our expert system, we conducted knowledge acquisition through structured and unstructured interviews with human experts, and we obtained heuristics for cost-effective water-utility operations. Analysis of these heuristics indicates that accurately estimating short-term (daily) forecasts of water demand is important so that energy-saving pumping schedules can be derived. However, the prediction of water demand is poorly understood—even by the experts, who approximate daily water demand based on their experience. The often inaccurate estimations result in inefficient operation of the water-dis-

tribution system. The lack of knowledge about water-demand prediction translates to a gap in the expert system's knowledge base. In other words, manual knowledge acquisition by itself is inadequate for handling all the situations that can arise in a complex engineering application.

The relevant studies on demand prediction have focused on developing mathematical models based on statistical studies of historical data.[1] In our project, we focused on using machine-learning techniques for generating rules on water-demand prediction, because mathematical modeling focuses on quantitative information alone. But in automated learning from observed data, both quantitative as well as qualitative variables are considered.

Because the database's training samples are incomplete and possibly ambiguous, exact decision rules cannot be derived by standard methods.[2–5] We propose a method for generating classification rules from incomplete information. This method extends the rough-set model,[6] and uses statistical information to define the positive and negative regions of a concept. Each classification rule generated by our learning system is characterized by an uncertainty factor, which is the probability that an object matching the condition part of the rule belongs to the concept.

Data collection and representation

We obtained data on the water domain from the municipal water department in Regina, Saskatchewan, Canada, and Environment Canada. The former provided us with historical data on water consumption, and the latter on weather conditions.

Characteristics of water demand. The instantaneous consumption of water in an urban distribution system depends on the many industrial, commercial, public, and domestic consumers distributed throughout the area supplied. Factors such as weather conditions, seasonal variation, day of the week, and whether a particular day is an observed holiday can all influence this consumption. We have found that expert operators base their water-demand predictions primarily on weather-related considerations and day of the week. Therefore, in this article, we use weather conditions and day of the week as the condition attributes, and consumer water demand as the decision attribute.

Condition attributes. Seven factors can affect the daily consumption of water in a city (see Table 1). The first factor is day of the week, which we chose because of the observation that on weekends the daily total-distribution flows are usually less than on weekdays. The city also has the power, through bylaws, to restrict watering of lawns on Wednesdays. Furthermore, Mondays are days of high water usage, because many people do their laundry on Mondays and because, in the summer, people water their lawns on the day after returning from a weekend at their cottages. The remaining factors are weather conditions: temperature, humidity, precipitation, and wind. We obtained the values of these factors from monthly meteorological summaries printed by Environment Canada.

Decision attribute. The total amount of water consumed by the city each day is calculated by summing the daily distribution flows metered at each pumping station of the city. The operator at the central control station records this number every day. The value varies from 50 megaliters in the coldest winter days to 180 megaliters in the summer.

Information systems. We assume that the given set of training samples represents the knowledge about the domain. In our approach, the training set is described by an *information system*.[4] The objects in a universe U are described by a set of attribute values.

Formally, an information system S is a quadruple $<U, A, V, f>$, where $U = \{x_1, x_2, \ldots, x_N\}$ is a finite set of objects, and A is a finite set of attributes. The attributes in A are further classified into two disjoint subsets: the *condition attributes* C and the *decision attribute* 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, where V_a is the *domain* of attribute a (the set of values of attribute a). $f: U \times A \rightarrow V$ is an information function that 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 306 training samples in our information system for water-demand prediction are objects that include daily information on the condition attributes and the decision attribute

Table 1. Condition factors for water-demand prediction.

LABEL	CONDITION ATTRIBUTES
a_0	Day of the week
a_1	Maximum temperature
a_2	Minimum temperature
a_3	Average humidity
a_4	Rainfall
a_5	Snowfall
a_6	Average wind speed

Table 2. Information system for water-demand prediction.

OBJECTS	a_0	a_1	a_2	a_3	a_4	a_5	a_6	D
obj_1	6	6	7	8	1	0	3	0
obj_2	1	5	7	9	0	0	1	0
obj_3	6	6	7	8	1	0	3	0
obj_4	3	6	7	4	0	0	1	1
obj_5	3	5	7	3	0	0	1	1
obj_6	3	6	7	4	0	0	1	1
obj_7	6	6	8	3	0	0	6	2
obj_8	3	6	7	4	0	0	1	2

(The C columns a_0 through a_6 are grouped under the header *C*.)

for 10 months, from March to December 1994. Table 2 lists eight of these objects. For the purpose of rough-set-based data analysis, we have generalized the information system by replacing the original attribute values with some discrete ranges—for example, we have discretized attribute a_2 (minimum temperature) into 10 categories, denoted as 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. For instance, 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. (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.)

The question of how to optimally discretize the attribute values is a subject of ongoing research; the method of discretization we have adopted is just one of many possibilities. Our experimental results are based on the whole set of training samples (the 306-object information system).

Data analysis

In what follows, we present the basics of an extended rough-set model, referred to as the probabilistic model of rough sets. The probabilistic model underlies the rule-extraction technique adopted in our research.

Indiscernibility relation. An information system provides only partial information of the universe. That is, the objects described

by the fixed set of selected attributes might be insufficient for characterizing the objects uniquely. Any two objects are indistinguishable from each other whenever they assume the same attribute values. This means we might not be able to distinguish all the objects solely by means of the admitted attributes and their values.

Given an information system $<U, A, V, f>$, 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 $R(B) = \{(x_i, x_j) \in U^2 \mid \forall a \in B, f(x_i, a) = f(x_j, a)\}$. We say that x_i and x_j are indiscernible to 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 information system shown in Table 2, obj_1 and obj_3 are indiscernible to the set of attributes $C \cup D$, and obj_4, obj_6, and obj_8 are indiscernible to the set of condition attributes C.

$R(B)$ is an equivalence relation on U for every $B \subseteq A$. Thus, we can define two natural equivalence relations, $R(C)$ and $R(D)$, on U for an information system S. A *concept* Y is an equivalence class of the relation $R(D)$. Without loss of generality, we can consider D as a singleton set. Our 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 $R(C)$:

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

We say that $x_i \in U$ *definitely belongs* to a concept Y if $[x_i]_R \subseteq Y$, and that $x_i \in U$ *possibly belongs* to the concept Y if $[x_i]_R \cap Y \neq \emptyset$.

We define conditional probabilities as

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

where $P(Y \mid [x_i]_R)$ is the probability of occurrence of event Y conditioned on event $[x_i]_R$. That is, $P(Y \mid [x_i]_R) = 1$ 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 1. Let us consider the given information system in Table 2. The concepts in this information system, the equivalence classes on the relation $R(D)$, are

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

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

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

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

The conditional probability of each equivalence class is

$P(Y_1 \mid X_1) = 0$
$P(Y_1 \mid X_2) = 0$
$P(Y_1 \mid X_3) = 2/3 = 0.667$
$P(Y_1 \mid X_4) = 1$
$P(Y_1 \mid X_5) = 0$

β-probabilistic approximation classification. Given an information system $S = \{U, A, V, f\}$ and an equivalence relation $R(C)$ (an *indiscernibility relation*) on U, an ordered pair $AS = <U, R(C)>$ is called an *approximation space*[4] 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 that are distinguishable in terms of the attributes and their values. Let $Y \subseteq U$ be a subset of objects representing a concept, and $R*(C) = \{X_1, X_2, ..., X_n\} = \{[x_1]_R, [x_2]_R, ..., [x_n]_R\}$ 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

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

and

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

The above definitions do not use the statistical information in the boundary region,

$$\overline{R(C)}(Y) - \underline{R(C)}(Y)$$

For this reason, several extensions to the original rough-set model have been proposed.[6–9,11] In our approach, we try to rectify this limitation by introducing a β-approximation space.

A β-approximation space AS_P is a triple $<U, R(C), P>$, where P is a probability measure and β is a real number in the range $(0.5, 1]$. The β-approximation space AS_P can be divided into the following regions:

β-positive region of the set Y:
$$POS_C(Y) = \bigcup_{P(Y|X_i) \geq \beta} \{X_i \in R*(C)\}$$

β-negative region of the set Y:
$$NEG_C(Y) = \bigcup_{P(Y|X_i) < \beta} \{X_i \in R*(C)\}$$

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

Let $x_i \in U$ be an object; $POS_C(Y)$ and $NEG_C(Y)$ are the positive and negative regions of the concept Y. The object x_i is classified as belonging to the concept Y if and only if $x_i \in POS_C(Y)$, or to the complement $\neg Y$ of the concept Y if and only if $x_i \in NEG_C(Y)$. We want 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 we deal with $POS_C(Y)$ and $NEG_C(Y)$ instead of the set Y. If $x_i \in U$ is in $POS_C(Y)$, it can be classified into the concept Y with the conditional probability $P(Y \mid X_i)$ greater than or equal to the parameter β.

Example 2. Following from Example 1, recall that $Y_1 = \{obj_4, obj_5, obj_6\}$. If we let $\beta = 1$, then the β-positive region of the set Y_1 is $POS_C(Y_1) = \{obj_5\}$, and the β-negative region of the set Y_1 is $NEG_C(Y_1) = \{obj_1, obj_2, obj_3, obj_4, obj_6, obj_7, obj_8\}$. If we let $\beta = 0.6$, then the β-positive region of the set Y_1 is $POS_C(Y_1) = \{obj_4, obj_5, obj_6, obj_8\}$, and the β-negative region of the set Y_1 is $NEG_C(Y_1) = \{obj_1, obj_2, obj_3, obj_7\}$.

Reduction of condition attributes. An information system often includes some condition attributes that do not provide any addi-

Objects	C a_1	a_3	Y_1	Number of objects in Y_1	Number of objects in $\neg Y_1$
obj_4, obj_6, obj_8	6	4	1	2	1
obj_5	5	3	1	1	0
obj_1, obj_3	6	8	0	0	2
obj_2	5	9	0	0	1
obj_7	6	3	0	0	1

tional information about the objects in U. Eliminating those attributes can reduce the complexity and cost of the decision process. We use the concept of a *reduct* in rough sets to describe the method of condition-attributes reduction.

Given an attribute-value system $S = <U, C \cup \{d\}, V, f>$, an attribute a is *dispensable* in C with respect to $\{d\}$ if $POS_{C-\{a\}}(Y) = POS_C(Y)$; otherwise a is an *indispensable* attribute in C with respect to $\{d\}$. A subset of condition attributes $B \subseteq C$ is a *dependent set* in S with respect to $\{d\}$ if a proper subset $K \subset B$ exists such that $POS_B(Y) = POS_K(Y)$; otherwise, B is an *independent set* with respect to $\{d\}$. A reduct C' of attributes C is a maximal independent subset of condition attributes with respect to $\{d\}$.[4]

The procedure for finding a single reduct is very straightforward. Consider a condition attribute $a \in C$. If the β-positive region $POS_{C-\{a\}}(Y)$ of the set Y is the same as $POS_C(Y)$, 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 can exist for a given attribute-value system. Selection of a *best* reduct depends on the optimality criterion associated with the attributes. We can also assign significance values to attributes and base the selection on those values.

Example 3. The condition attributes C of the information system in Table 2 has a total of seven reducts. Table 3 shows a reduced information system based on the reduct $\{a_1, a_3\}$ with respect to the concept Y_1. The objects with value 1 for attribute Y_1 belong to the β-positive region of the concept Y_1; the objects with value 0 for attribute Y_1 belong to the β-negative region of concept Y_1. Here, $\beta = 0.6$.

Generating decision rules

Rule generation is a crucial task in any learning system. We now describe how decision rules are generated based on the reduct obtained in the previous section.

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 that 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 called the *description* of the equivalence class $X_i \in R^*(RED)$. We can express the description of X_i as

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

where $\wedge$ 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 and $x_i \in Y$.

The following decision rules describe the relationship between the partition $R^*(RED)$ and the partition $R^*(D)$:

for $X_i \in R^*(RED)$,

$$Des(X_i) \xrightarrow{c_i} Des(Y), \text{ if } P(Y|X_i) \geq \beta$$

$$Des(X_i) \xrightarrow{c_i} Des(\neg Y), \text{ if } P(Y|X_i) < \beta$$

where c_i is the uncertainty 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 definitely belongs to Y with uncertainty c_i. Similarly, if $P(Y|X_i) < \beta$, then the object x_i possibly belongs to the complementary concept $\neg Y$ with uncertainty c_i.

Example 4. Following from Example 3, let RED denote the reduct $\{a_1, a_3\}$. The collection $R^*(RED)$ of the equivalence classes of the relation $R(RED)$ is

$R^*(RED) = \{X_1, X_2, X_3, X_4, X_5\} = \{\{obj_4, obj_6, obj_8\}, \{obj_5\}, \{obj_1, obj_3\}, \{obj_2\}, \{obj_7\}\}$

The descriptions of these equivalence classes are

$Des(X_1) = (a_1 = 6) \wedge (a_3 = 4)$
$Des(X_2) = (a_1 = 5) \wedge (a_3 = 3)$
$Des(X_3) = (a_1 = 6) \wedge (a_3 = 8)$
$Des(X_4) = (a_1 = 5) \wedge (a_3 = 9)$
$Des(X_5) = (a_1 = 6) \wedge (a_3 = 3)$

The descriptions of Y_1 and $\neg Y_1$ are

$Des(Y_1) = (D = 1)$
$Des(\neg Y_1) = (D \neq 1)$

Because $Y_1 = \{obj_4, obj_5, obj_6\}$, we calculate the condition probabilities as

$P(Y_1 | X_1) = 2/3 = 0.67$
$P(Y_1 | X_2) = 1$
$P(Y_1 | X_3) = 0$
$P(Y_1 | X_4) = 0$
$P(Y_1 | X_4) = 0$

We then obtain the decision rules with respect to the concept Y_1 as follows:

r_1^+: $(a_1 = 6) \wedge (a_3 = 4) \xrightarrow{0.67} (D = 1)$
r_2^+: $(a_1 = 5) \wedge (a_3 = 3) \xrightarrow{1} (D = 1)$
r_1^-: $(a_1 = 6) \wedge (a_3 = 8) \xrightarrow{1} (D \neq 1)$
r_2^-: $(a_1 = 5) \wedge (a_3 = 9) \xrightarrow{1} (D \neq 1)$
r_3^-: $(a_1 = 6) \wedge (a_3 = 3) \xrightarrow{1} (D \neq 1)$

Rule generalization. As indicated in the previous subsection, we can obtain decision rules directly from the reduced information system. We obtain one rule for each equivalence class of the partition $R^*(RED)$. However, rules might contain attributes whose values are irrelevant for determining the target concept. Furthermore, we can generalize rules further 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, maximally general rules contain a minimum number of conditions. We use the *decision matrix* technique to find all the maximally general rules.[10]

Let X_i^+, $i = (1, 2, \ldots \gamma)$, denote the equiva-

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

	X_1^-	X_2^-	X_3^-
X_1^+	$(a_3, 4)$	$(a_1, 6), (a_3, 4)$	$(a_3, 4)$
X_2^+	$(a_1, 5), (a_3, 3)$	$(a_3, 3)$	$(a_1, 5)$

lence classes of the relation $R^*(\text{RED})$ such that $X_i^+ \subseteq \text{POS}_{\text{RED}}(Y)$; and let X_j^-, $j = (1, 2, \ldots, \rho)$, denote the equivalence classes of the relation $R^*(\text{RED})$ such that $X_j^- \subseteq \text{NEG}_{\text{RED}}(Y)$. We define a decision matrix $M = (M_{ij})_{\gamma \times \rho}$ as

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

where a is a condition attribute belonging to RED. That is, 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^-.

We obtain the set of decision rules computed for a given equivalence class X_i^+ by treating each element of M_{ij} as a Boolean expression and constructing the following Boolean function:

$$B_i = \bigwedge_j \left(\vee M_{ij} \right)$$

where $\wedge$ and $\vee$ are the usual conjunction and disjunction operators.

We can show that the prime implicants of the Boolean function B_i are the maximally general rules for the equivalence class X_i^+ belonging to the positive learning region $\text{POS}_{\text{RED}}(Y)$. Thus, by finding the prime implicants of all the decision functions B_i ($i = 1, 2, \ldots, \gamma$), we can compute all the maximally general rules for the positive learning region $\text{POS}_{\text{RED}}(Y)$.

Example 5. Following from Example 4, the equivalence classes of the relation $R(\text{RED})$ are

$$X_1^+ = X_1 = \{obj_4, obj_6, obj_8\}$$
$$X_2^+ = X_2 = \{obj_5\}$$
$$X_1^- = X_3 = \{obj_1, obj_3\}$$
$$X_2^- = X_4 = \{obj_2\}$$
$$X_3^- = X_5 = \{obj_7\}$$

Table 4 shows the decision matrix

$$M = \left(M_{ij} \right)_{2 \times 3}$$

We then obtain the Boolean functions for each X_i^+ ($i = 1, 2$) as follows:

$$B_1 = (a_3 = 4) \wedge ((a_1 = 6) \vee (a_3 = 4))$$
$$\wedge (a_3 = 4) = (a_3 = 4)$$
$$B_2 = ((a_1 = 5) \vee (a_3 = 3)) \wedge (a_3 = 3)$$
$$\wedge (a_1 = 5) = (a_1 = 5) \wedge (a_3 = 3)$$

Therefore, the maximally general rule for the equivalence class X_1^+ is

$$(a_3 = 4) \rightarrow^{0.67} (D = 1)$$

and the maximally general rule for the equivalence class X_2^+ is

$$(a_1 = 5) \wedge (a_3 = 3) \rightarrow^1 (D = 1).$$

Given the set of all maximally general rules for an information system S, our system provides the options to find the set of *minimal rules* and the set of *minimal covering rules*. The support set of a rule r_i, denoted as $\text{supp}(r_i)$, is the collection of rows of the original table satisfying the condition part of r_i. A collection of rules RUL′ is a set of minimal rules if, for every $r_i \in$ RUL′,

$$\text{supp}(r_i) \not\subseteq \text{supp}(r_j)_{\forall r_j \in \text{RUL}', r_i \neq r_j}$$

A collection of rules RUL″ is a set of minimal covering rules if, for every $r_i \in$ RUL″,

$$\text{supp}(r_i) \not\subseteq \bigcup \text{supp}(r_j)_{r_j \in \text{RUL}'', r_i \neq r_j}$$

Example 6. As shown in Example 5, we can use a decision matrix to calculate the maximally general rule for the concept Y_1. Similarly, we can obtain the maximally general rules for concepts Y_0 and Y_2. The set of all maximally general rules for the information system in Table 2 are

$$r_1: (a_3 = 8) \rightarrow^1 (D = 0)$$
$$r_2: (a_3 = 9) \rightarrow^1 (D = 0)$$
$$r_3: (a_3 = 4) \rightarrow^{0.67} (D = 1)$$
$$r_4: (a_1 = 5) \wedge (a_3 = 3) \rightarrow^1 (D = 1)$$
$$r_5: (a_1 = 6) \wedge (a_3 = 3) \rightarrow^1 (D = 2)$$

where r_1 and r_2 are for the concept Y_0, r_3 and r_4 are for the concept Y_1, and r_5 is for the concept Y_2.

The support sets of these rules are

$$\text{supp}(r_1) = \{obj_1, obj_3\}$$
$$\text{supp}(r_2) = \{obj_2\}$$
$$\text{supp}(r_3) = \{obj_4, obj_6, obj_8\}$$
$$\text{supp}(r_4) = \{obj_5\}$$
$$\text{supp}(r_5) = \{obj_7\}$$

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

Experimental results

We have implemented this method in the KDD-R (*K*nowledge *D*iscovery in *D*atabases, *R*ough Sets Approach) system developed at the University of Regina.[9] The system performs data analysis, database mining, pattern recognition and validation, and expert-system building. We tested the proposed method by applying the KDD-R system on the data for water-demand prediction. Our objective is to analyze the set of training data and generate a set of decision rules to predict a city's daily water demand. As we mentioned, the set of training samples consists of 306 objects collected over a period of 10 months and includes information on day of the week, weather conditions, and daily water consumption.

In our experiment, we divided the values of decision attributes into nine ranges so that the information system has nine concepts. The KDD-R system generated a total of 188 rules. Table 5 lists the number of rules generated for different concepts; *NC* stands for the number of training samples covered by the rules.

The most general rule for the concept $D = (60 - 70]$ is

$$(-13.80 < a_2 \leq -1.40) \wedge (81.60 < a_3 \leq 93.00) \rightarrow^1 (60 < D \leq 70)$$

This rule covers 12.37% of the training objects concluding the concept. The rule states that

If the maximum temperature is within the range (−13.80°C, −1.40°C] and the average humidity is within (81.60%, 93%], then the water demand is between 60 megaliters and 70 megaliters, with an uncertainty factor of 1.

The most general rule for the concept $D = (70 - 80]$ is

$$(a_0 = M) \wedge (-31.50 \leq a_2 \leq -13.80) \rightarrow^1 (70 < D \leq 80)$$

This rule covers 4.17% of the training objects concluding the concept. It states that

If the day of the week is Monday and the minimum temperature is within the range [–31.50°C, –13.80°C], then the water demand is between 70 megaliters and 80 megaliters, with an uncertainty factor of 1.

The most general rule for the concept $D = (80 – 90]$ is

$(a_0 = F) \land (6.70 < a_1 \leq 34.10) \land (4.50 \leq a_6 \leq 8.70) \rightarrow {}^1 (80 < D \leq 90)$

This rule covers 7.90% of the training objects concluding the concept. It states that

If the day of the week is Friday and the maximum temperature is within the range (6.70°C, 34.10°C] and the average wind speed is within [4.50 km/hr., 8.70 km/hr.], then the water demand is between 80 megaliters and 90 megaliters, with an uncertainty factor of 1.

The most general rule for the concept $D = (100 – 110]$ is

$(a_0 > SU) \land (47.40 < a_3 \leq 53.10) \land (17.10 < a_6 \leq 21.30) \rightarrow {}^1 (100 < D \leq 110)$

This rule covers 22.22% of the training objects concluding the concept. It states that

If the day of the week is Sunday and the average humidity is within (47.40%, 53.10%] and the average wind speed is within (17.10 km/hr., 21.30 km/hr.], then the water demand is between 100 megaliters and 110 megaliters, with an uncertainty factor of 1.

To evaluate the rules derived by our method, we conducted a *leave-ten-out* experiment by using 90% of the data for training and the remaining 10% for testing. The error rate depends on the selection of training samples. We conducted the experiment 10 times. The best error rate of prediction was 6.67%, and the average error rate of prediction was 10.27%.

WE HAVE SUGGESTED A METHOD for generating prediction rules from a given set of training examples. The method is an extension of the rough-set model. The salient feature of the proposed method is that it uses the statistical information inherent in the

Table 5. Number of rules for different concepts.

CONCEPTS	NUMBER OF RULES	NC
[53 – 60]	3	3
(60 – 70]	43	97
(70 – 80]	67	120
(80 – 90]	32	38
(90 – 100]	16	18
(100 – 110]	8	9
(110 – 120]	6	7
(120 – 130]	4	5
(130 – 140]	3	3
(140 – 176]	5	6

knowledge system. Thus, our method can derive decision rules from incomplete knowledge. This capability is important, because we seldom have complete and consistent information when designing intelligent systems.

Application of this knowledge-discovery method for water-demand prediction complements manual knowledge-acquisition techniques (see the sidebar, "The intelligent system"). It offers the advantage of describing important relationships between condition factors and water consumption in terms of simple if-then rules that users can easily understand. The experimental results indicate that the proposed algorithm can generate rules for water-demand prediction, providing more precise information than is available through knowledge acquisition from human experts.

Our proposed method for machine induction is general and can be applied to other domains, such as general consumer-demand prediction, fault diagnosis, and process control. For water-demand prediction, we have also tried different sets of causal variables with the rule-induction method to further improve predictive accuracy. The results show that adding time-series data, such as yesterday's and the day before yesterday's water consumption, as conditional variables gives better results than using weather factors alone. In terms of the

The intelligent system

The knowledge-discovery module for water-demand prediction is one component of the intelligent system for monitoring and controlling the water-distribution system. The intelligent system consists of five modules:

- the expert-system module,
- the energy-management module,
- the water-demand prediction module,
- the pipeline-network-simulation module, and
- the scheduling and planning module.

The expert system consists of a knowledge base for detecting faults and recommending a series of adjustments on equipment at a given state of operations.

The energy-management module informs users of the most cost-effective combination of pumps and valves among the possible configurations suggested by the expert system.

The prediction module forecasts water demands. It has been implemented using various AI techniques, including neural networks, fuzzy sets, case-based reasoning, and knowledge discovery from databases.

The work presented in the main text of the article describes the knowledge-discovery component of the prediction module. The forecast on water usage, along with the recommendations on pumps and valve adjustments provided by the expert system, becomes input to the simulation module, for verification purposes.

The simulation program uses an extended period simulation to represent the dynamic behavior of flows, pressures, and water levels. The operational procedure of pumps and valves suggested by the expert system are satisfactory if the simulation module shows that

- the pressure at the outlet of each pumping station is within the set-point limit,
- the water level remains within the range allowable for the reservoirs, and
- the combination of pumps and valves does not cause any reservoir to overflow or deplete within a short period of time.

The project began in May 1994, and we have completed the prototype of the expert system, the simulation program, the prediction module, and the energy-management module. We are working on the scheduling and planning module, and on integrating the different components into an intelligent system. We expect to complete the project this fall.

machine-induction method, we have developed a new algorithm that integrates rule induction with case-based reasoning to improve case retrieval and problem solving. We have applied this integrated method to numeric-prediction domains. Experimental results show that case-based reasoning with rule induction performs better than case-based reasoning by itself, in terms of predictive accuracy.

Acknowledgments

References

1. J. Quevedo et al., "Time Series Modelling of Water Demand: A Study on Short-Term and Long-Term Predictions," in *Computer Applications in Water Supply*, Vol. 1, B. Coulbeck and C. Orr, eds., Research Studies Press Ltd., Letchworth, England, 1988, pp. 268–288.

2. R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, eds., *Machine Learning: An Artificial Intelligence Approach,* Vol. 1, Morgan Kaufmann, San Francisco, 1983.

3. R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, eds., *Machine Learning: An Artificial Intelligence Approach,* Vol. 2, Morgan Kaufmann, 1986.

4. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Norwell, Mass., 1991.

5. J.R. Quinlan, "Induction of Decision Trees," *Machine Learning*, Vol. 1, No. 1, 1986, pp. 81–106.

6. W. Ziarko, "Variable Precision Rough Set Model," *J. Computer and System Sciences*, Vol. 46, No. 1, 1993, pp. 39–59.

7. J. Katzberg and W. Ziarko, "Variable Precision Rough Sets with Asymmetric Bounds," in *Rough Sets, Fuzzy Sets and Knowledge Discovery*, W. Ziarko, ed., Springer-Verlag, New York, 1994, pp. 167–177.

8. Z. Pawlak, S.K.M. Wong, and W. Ziarko, "Rough Sets: Probabilistic Versus Deterministic Approach," *Int'l J. Man-Machine Studies*, Vol. 29, No. 1, 1988, pp. 81–95.

9. W. Ziarko and N. Shan, "KDD-R: A Comprehensive System for Knowledge Discovery in Databases Using Rough Sets," *Proc. Third Int'l Workshop Rough Sets and Soft Computing*, San Jose State Univ., San Jose, Calif., 1994, pp. 164–173.

10. W. Ziarko and N. Shan, "A Rough Set-Based Method for Computing All Minimal Deterministic Rules in Attribute-Value Systems," *Computational Intelligence: An Int'l J.*, Vol. 12, No. 2, 1996, pp. 223–234.

11. S.K.M. Wong and W. Ziarko, "INFER—An Adoptive Decision Support System Based on the Probabilistic Approximate Classification," *Proc. Sixth Int'l Workshop Expert Systems and Their Applications*, Agence de Informatique, Avignon, France, 1986, pp. 713–726.

Aijun An is a PhD candidate in the Department of Computer Science at the University of Regina. Her research interests include machine learning, case-based reasoning, knowledge engineering for developing real-time expert systems, validation and verification of knowledge-based systems, and applications of artificial intelligence techniques to practical problems. She received a BSc and an MSc in computer science from Xidian University, Xi'an, China. She is a student member of the AAAI and the IEEE Computer Society. Readers can contact An at the Dept. of Computer Science, Univ. of Regina, Regina, SK S4S 0A2, Canada; aijun@cs.uregina.ca.

Christine Chan is an associate professor in the Department of Computer Science at the University of Regina. Her research interests include industrial applications of artificial intelligence, object-oriented methodologies for knowledge-based systems development, knowledge acquisition and conceptual modeling, and educational instructional software. She received MSc degrees in computer science and management information systems from the University of British Columbia. She also received a PhD in the interdisciplinary studies of computer science, philosophy, and psychology from Simon Fraser University. She is a member of the AAAI, the IEEE Computer Society, the ACM, and the Canadian Information Processing Society. Readers can contact Chan at the Dept. of Computer Science, Univ. of Regina, Regina, SK S4S 0A2, Canada; chan@cs.uregina.ca; http://www.cs.uregina.ca/~chan.

Ning Shan is a PhD candidate in the Department of 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 from ChengDu Electronic Engineering Institute, China, and an MSc in computer science from the University of Regina. Readers can contact Shan at the Dept. of Computer Science, Univ. of Regina, Regina, SK S4S 0A2, Canada; ning@cs.uregina.ca.

Nick Cercone is a professor and the chair of the Computer Science Department at the University of Waterloo. His research interests include natural-language processing, knowledge-based systems, knowledge discovery in databases, and design and human interfaces. He received a BS in engineering science from the University of Steubenville, an MS in computer and information science from Ohio State University, and a PhD in computing science from the University of Alberta. He is a member of the ACM, the IEEE, the AAAI, the Artificial Intelligence Simulation of Behavior, the AGS, and the Association of Computational Linguistics. He recently won the Canadian Artificial Intelligence Society's Distinguished Service Award. Readers can contact Cercone at the Computer Science Dept., Faculty of Mathematics, Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada; ncercone@uwaterloo.ca.

Wojciech Ziarko is a professor in the Computer Science Department at the University of Regina. His research interests are knowledge discovery in databases, machine learning, pattern classification, and control-algorithm acquisition from sensor 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 mathematical framework in much of his work. He received an MSc in applied mathematics from Warsaw University of Technology, Poland, and a PhD in computer science, specializing in databases, from the Institute of Computer Science of the Polish Academy of Sciences, Warsaw, Poland. Readers can contact Ziarko at the Dept. of Computer Science, Univ. of Regina, Regina, SK S4S 0A2, Canada; ziarko@cs.uregina.ca.