Yiyu Yao and Noroozi
A unified model for set-based computations,
Proceedings of the Third International Workshop on Rough Sets and Soft Computing,
T.Y. Lin and A.M. Wildberger (Eds.), San Jose State University, pp. 252-255, 1995.

A UNIFIED MODEL FOR SET-BASED COMPUTATIONS

Y.Y. Yao and N. Noroozi

Department of Mathematical Sciences, Lakehead University
Thunder Bay, Ontario, Canada P7B 5E1, e-mail: yyao@thunder.lakeheadu.ca

Computational methods, information systems, uncertainty management.

ABSTRACT

This paper examines a conceptual model for set-based computations, which extends point-based operations and relations into their set-based counterparts. Interval-number algebra, interval-set algebra, and set-based information systems may be viewed as special cases. It is argued that the proposed model may be useful in establishing a unified framework for uncertainty representation and approximate reasoning.

1 INTRODUCTION

Extensive studies on uncertainty management have resulted in many useful extensions of classical mathematical tools. A class of such tools is the interval-based computations, which use a pair of lower and upper bounds to specify the range of an ill-known value. Examples of interval-based approaches are interval-number algebra, interval-set algebra, interval algebra, rough sets, incidence calculus, fuzzy sets, belief and plausibility functions, necessity and possibility functions, interval-valued probabilistic reasoning, and interval-valued fuzzy reasoning (Dubois and Prade 1988; Moore 1966; Pawlak 1982; Ruspini 1991; Shafer 1976; Yao 1993).

One of the motivations of interval-based computations is that in many practical situations it is impossible to specify the exact values of certain parameters under consideration. One may only be able to provide a range within which the true value lies. Results from earlier studies suggest that a number of important issues should be considered. It is necessary to establish a unified framework as a basis for the study of interval-based computation methods. Such a framework may provide a common ground for various approaches using different notations. An underlying assumption of interval-based computations is that intervals are defined on ordered sets. In many cases, an order relation may not exist on the values of certain parameters. By observing the fact that an interval indeed specifies a set of values, it is natural to generalize interval-based computations into set-based computations.

The main objective of this paper is to investigate basic issues of set-based computations. Operations on set values are defined based on the corresponding point-based (i.e., single-value-based) operations on their members. The properties of set-based computations are examined in connection to the corresponding properties of the point-based computations. The results of this study will be useful in establishing a framework for more systematic study of set-based computations.

2 SET-BASED COMPUTATIONS

Various types of computational methods can be modeled by different mathematical systems. For example, in numerical computations, one uses the numerical system $(\Re, +, -, \times, /, \geq, =)$, where $\Re$ is the set of real numbers, $+, -, \times$ and $/$ are binary operations on $\Re$, and $\geq$ and $=$ are binary relations on $\Re$. In logical inference, one uses the system $(V, \wedge, \vee, \neg)$, where V is a set of truth values, and $\wedge$, $\vee$ and $\neg$ are operations defined on V. Abstracting from these examples, we adopt the concept of relational systems (Roberts 1979). A relational system is an ordered $(p+q+1)$-tuple,

$$RS = (U, \circ_1, \ldots, \circ_p, R_1, \ldots, R_q), \tag{1}$$

where U is a nonempty set, $\circ_1, \ldots, \circ_p$ are operations on U, and $R_1, \ldots, R_q$ are (not necessarily binary) relations on U. In this study, we only consider unary and binary operations, and binary relations.

In a relational system, all the operations and relations are defined on elements of the set U. When it is impossible to represent a physical quantity using a single element of U, a subset of U may be used. To accommodate this set-based representation scheme, one has to extend the operations and relations on the elements of U into operations and relations on the subsets of U. We refer to the former as point-based computations, and the latter as set-based computations. The extended relational system must be defined in such a way that it preserves important characteristics of the original system. In particular, if singleton sets of U are used, set-based computations must reduce to point-based computations.

A binary operation $\circ$ on U can be extended into a binary operation $\circ'$ on 2^U by applying it to the members of the subsets of U. Given two subsets $A, B \in 2^U$, one can derive another subset $C \in 2^U$ by collecting all elements $a \circ b$, where $a \in A$ and $b \in B$.

Definition 1. Suppose $\circ$ is a binary operation on U. An extended binary operation $\circ'$ on $2^U - \{\emptyset\}$ is defined by:

$$A \ \circ' \ B = \{a \circ b \mid a \in A, b \in B\}. \qquad (2)$$

If only singleton subsets of U are used, operation $\circ'$ reduces to $\circ$. By definition, many properties of operations in the original system can be carried over by the extended operations.

Theorem 1. Suppose $\circ$ is a binary operation on U, and $\circ'$ on $2^U - \{\emptyset\}$ is the extended binary operation defined by equation (**??**). Then,

(a). if $\circ$ is commutative, $\circ'$ is commutative,
(b). if $\circ$ is associative, $\circ'$ is associative.

In general, it is not necessary to use the entire set 2^U to construct a system for set-based computations. In many situations, one may find that it is more meaningful to use a subset of 2^U. For example, if U is an ordered set, one may consider the set of all closed intervals, which is only a subset of 2^U.

The extension of a binary relation R can be done in a similar manner. However, the process is more complicated because the relation may only hold for some elements of two subsets of U. Given two subsets $A, B \in 2^U - \{\emptyset\}$, relation R may not hold for any pair (a, b), where $a \in A$ and $b \in B$. In the case when the relation R does hold between elements of A and B, it may hold for only one pair, two pairs, ..., or all pairs. We may, for example, define a family of graded relations based on the number of pairs for which the relation holds. In the present study, we consider the two extreme points of this spectrum.

Definition 2. Suppose R is a binary relation on U. A pair of extended binary relations $\langle R_*, R^* \rangle$ on $2^U - \{\emptyset\}$ is defined by:

$$\begin{aligned} A \ R_* \ B &\iff (\forall a \in A, \forall b \in B) \ aRb, \\ A \ R^* \ B &\iff (\exists a \in A, \exists b \in B) \ aRb. \end{aligned} \qquad (3)$$

These two relations can be interpreted as representing necessity and possibility. If $A \ R_* \ B$, an element of A is necessarily related to an element of B. If $A \ R^* \ B$, an element of A is possibly related to an element of B. It can be seen that the following relationship holds: for $A, B \in 2^U - \{\emptyset\}$,

$$A \ R_* \ B \Longrightarrow A \ R^* B. \qquad (4)$$

That is, if a pair of elements from two sets are necessarily related, they are possibly related. If only singleton subsets of U are used, both R_* and R^* reduce to relation R.

Theorem 2. Suppose R is a binary relation on U, and $\langle R_*, R^* \rangle$ on $2^U - \{\emptyset\}$ is a pair of extended binary relations defined by equation (**??**). Then,

(a). if R is reflexive, R^* is reflexive,
(b). if R is symmetric, R_* and R^* are symmetric,
(c). if R is transitive, R_* is transitive.

From Definitions 1 and 2, given a relational system

$$RS = (U, \circ_1, \ldots, \circ_p, R_1, \ldots, R_q),$$

we can extend it into a set-based system:

$$RS' = (\Gamma(2^U), \circ'_1, \ldots, \circ'_p, \langle R_{1*}, {R_1}^* \rangle, \ldots, \langle R_{q*}, {R_q}^* \rangle).$$

The set $\Gamma(2^U) \subseteq 2^U$ should be chosen such that the extended operations $\circ'$ are closed. System RS' is a natural generalization of pointed-based system RS. It should be noted that we have not put any constraints on the elements of U. Elements of U may in fact be sets themselves. Thus, the proposed framework provides a simple, yet powerful enough, model of set-based computations.

To illustrate the usefulness the proposed framework, the following examples analyze briefly a few existing set-based computation approaches. It will be shown that all these models conform to the same basic structure introduced in this section.

Example 1. Interval-number algebra. An *interval number* $[a_1, a_2] = \{x \mid a_1 \leq x \leq a_2\}$, where $a_1 \leq a_2$, is an interval of real numbers. a_1 is referred to as the lower bound and a_2 the upper bound. Degenerate intervals of the form $[a, a]$ are equivalent to real numbers. Let I and J be two interval numbers, and let $*$ denote an arithmetic operation $+$, $-$, $\times$ or $/$ on pairs of real numbers. By Definition 1, an arithmetic operation $*$ may be extended to pairs of interval numbers I, J by:

$$I * J = \{x * y \mid x \in I, y \in J\}.$$

The result $I * J$ is again a closed bounded interval unless $0 \in J$ and the operation $*$ is division (in which case, $I * J$ is undefined). It can be easily verified that many properties of arithmetic operations can be carried over to interval-number operations. For example, interval-number addition is commutative and associative.

Using Definition 2, we can define two binary relations $\geq_*$ and $\geq^*$ on interval numbers based on $\geq$:

$$\begin{aligned} I \geq_* J &\Longleftrightarrow \text{for all } x \in I \text{ and } y \in J, \quad x \geq y; \\ I \geq^* J &\Longleftrightarrow \text{there exist a } x \in I \\ &\qquad \text{and a } y \in J \text{ such that } x \geq y. \end{aligned}$$

The relation $\geq_*$ is not reflexive but transitive, whereas $\geq^*$ is reflexive but not transitive. One may similarly define two relations $=_*$ and $=^*$ based on the equality relation $=$ on the real numbers. The arithmetic system defined above is called *interval arithmetic* or an *interval-number algebra*.

Example 2. Interval-set algebra. Let U be a finite nonempty set called the universe and 2^U its power set. Given two sets $A_1, A_2 \in 2^U$ with $A_1 \subseteq A_2$, the following subset of 2^U,

$$\mathcal{A} = [A_1, A_2] = \{X \in 2^U \mid A_1 \subseteq X \subseteq A_2\},$$

is called a closed *interval set*. The set A_1 is called the lower bound of the interval set and A_2 the upper bound. Let $\cap, \cup$ and $-$ denote set intersection, union and difference, respectively. From Definition 1, we have the corresponding binary operations on interval sets: For $\mathcal{A} = [A_1, A_2]$ and $\mathcal{B} = [B_1, B_2]$,

$$\begin{aligned} \mathcal{A} \sqcap \mathcal{B} &= \{X \cap Y \mid X \in \mathcal{A}, Y \in \mathcal{B}\}, \\ \mathcal{A} \sqcup \mathcal{B} &= \{X \cup Y \mid X \in \mathcal{A}, Y \in \mathcal{B}\}, \\ \mathcal{A} \setminus \mathcal{B} &= \{X - Y \mid X \in \mathcal{A}, Y \in \mathcal{B}\}. \end{aligned}$$

Let $I(2^U)$ denote the set of all closed interval sets. Then the above defined operations are closed on $I(2^U)$, namely, $\mathcal{A} \sqcap \mathcal{B}$, $\mathcal{A} \sqcup \mathcal{B}$ and $\mathcal{A} \setminus \mathcal{B}$ are interval sets. The interval-set complement $\neg[A_1, A_2]$ of $[A_1, A_2]$ is defined by $[U, U] \setminus [A_1, A_2]$. This is equivalent to $[U - A_2, U - A_1] = [A_2^c, A_1^c]$, where $A^c = U - A$ denotes set complement of A. Clearly, $\neg[\emptyset, \emptyset] = [U, U]$ and $\neg[U, U] = [\emptyset, \emptyset]$. Degenerate interval sets of the form $[A, A]$ are equivalent to ordinary sets. For degenerate interval sets, the proposed operations $\sqcap$, $\sqcup$, $\setminus$, and $\neg$ reduce to the usual set-theoretic operations. Properties of interval-set operations can be derived from the properties of set operations. For instance, the idempotent, commutativity, associativity, distributivity, absorption, De Morgan's laws and double negation law hold. However, for an interval set $\mathcal{A}$, $\mathcal{A} \sqcap \neg\mathcal{A}$ is not necessarily equal to $[\emptyset, \emptyset]$, $\mathcal{A} \sqcup \neg\mathcal{A}$ is not necessarily equal to $[U, U]$, and $\mathcal{A} \setminus \mathcal{A}$ is not necessarily equal to $[\emptyset, \emptyset]$. Therefore, $I(2^U)$ is a completely distributive lattice but not a Boolean algebra.

From Definition 2, we can derive two binary relations $\supseteq_*$ and $\supseteq^*$ on interval sets:

$$\begin{aligned} \mathcal{A} \supseteq_* \mathcal{B} &\Longleftrightarrow \text{for all } X \in \mathcal{A} \text{ and } Y \in \mathcal{B}, \quad X \supseteq Y; \\ \mathcal{A} \supseteq^* \mathcal{B} &\Longleftrightarrow \text{there exist a } X \in \mathcal{A} \\ &\qquad \text{and a } Y \in \mathcal{B} \text{ such that } X \supseteq Y. \end{aligned}$$

Relation $\supseteq_*$ is not reflexive but transitive, whereas $\supseteq^*$ is reflexive but not transitive.

Example 3. Set-based information systems. We define a set-based information system to be a quadruple, following Lipski (1981), Pawlak (1981), Vakarelov (1991):

$$S = (O, A, \{V_a \mid a \in A\}, \{f_a \mid a \in A\}),$$

where

O is a nonempty set of objects,

A is a nonempty set of attributes,

V_a is a nonempty set of values of $a \in A$,

$f_a : O \times A \longrightarrow 2^{V_a}$ is an information function.

If all information functions map an object only to singleton sets of attribute values, we obtain a degenerate set-based information system commonly used in the rough-set model (Pawlak 1982).

Set-based computations introduced in this section can be easily applied in the set-based information systems. We use the following information system to demonstrate the main idea.

	Age	Height	Language
o_1	{35}	{tall}	{English, French}
o_2	[30, 35]	{medium}	{French}
o_3	{20}	[medium, tall]	{English}
o_4	{60, 61}	{short}	{English, French}
o_5	{54}	[short, tall]	{English}

Suppose all attributes take a single value and a set-based information system is used to represent uncertainty in specifying the actual value. For an unordered set of attribute values, a valid operation is the comparison of equality. Consider a query q_1 : Language = English. In the case, we have two possible retrieved sets derived from $=_*$ and $=^*$:

$$Ret_*(q_1) = \{o_3, o_5\}, \quad Ret^*(q_1) = \{o_1, o_3, o_4, o_5\}.$$

Elements of Ret_* definitely satisfy the query, whereas elements of $Ret^* - Ret_*$ may satisfy the query. The pair (Ret_*, Ret^*) defines an interval set $[Ret_*, Ret^*]$, indicating the range of the set of objects that actually satisfies the query. For an ordered set of attribute values, in addition to the comparison of equality, we may also make comparisons using the order relation. For a given order relation $\succeq$ defined on V_a, it induces two extended relations $\succeq_*$ and $\succeq^*$. Accordingly, two sets Ret_* and Ret^* will be produced. For example, for the query, q_2 : Age $\geq$ 34, we have:

$$Ret_*(q_2) = \{o_1, o_4, o_5\}, \quad Ret^*(q_2) = \{o_1, o_2, o_4, o_5\}.$$

They are a pair of lower and upper approximations.

Combining q_1 and q_2, we obtain two composite queries: q_1 **and** q_2, q_1 **or** q_2. Using these queries, we derive the following two pairs of retrieved sets:

$$\begin{aligned}
&Ret_*(q_1 \textbf{ and } q_2) = \{o_5\},\\
&Ret^*(q_1 \textbf{ and } q_2) = \{o_1, o_4, o_5\};\\
&Ret_*(q_1 \textbf{ or } q_2) = \{o_1, o_3, o_4, o_5\},\\
&Ret^*(q_1 \textbf{ or } q_2) = \{o_1, o_2, o_3, o_4, o_5\}.
\end{aligned}$$

Obviously, the following properties hold:

$$\begin{aligned}
Ret_*(q_1 \textbf{ and } q_2) &= Ret_*(q_1) \cap Ret_*(q_2),\\
Ret^*(q_1 \textbf{ and } q_2) &= Ret^*(q_1) \cap Ret^*(q_2),\\
Ret_*(q_1 \textbf{ or } q_2) &= Ret_*(q_1) \cup Ret_*(q_2),\\
Ret^*(q_1 \textbf{ or } q_2) &= Ret^*(q_1) \cup Ret^*(q_2).
\end{aligned}$$

They correspond to the operations of the interval-set algebra. These rules may be considered as a generalization of the rules used in database systems. They may be used in a retrieval process of a set-based information systems. However, it should be noted that they may not generate the tightest bounds.

3 CONCLUSION

In this paper, we have presented and analyzed a framework for set-based computations. This framework is particularly useful in situations where it is difficult to obtain a precise value of certain parameter, or where set-valued attributes play an important role. To be consistent with point-based computations, operations on set values are defined by carrying over the corresponding point-based operations. A number of existing set-based computation methods, such as interval-number algebra, interval-set algebra, and set-valued information systems, are special cases.

This paper focused mainly on the conceptual unification of existing approaches of set-based computations. Such an integration of various models is a preliminary and an important step towards the establishment of a unified framework. The definition of the extended system suggested in this paper represents only one of many possible ways. It is entirely possible that a different definition may be more appropriate in different applications. For example, one can mix the quantifiers in Definition 2 to obtain additional extended relations. In fact, with respect to the interval-set operations, similar and alternative proposals have been suggested by many authors (Calabrese 1987; Goodman *et al.* 1991; Wellman and Simmons 1988). It is worthwhile to incorporate the results of these studies into the proposed framework.

REFERENCES

Calabrese, P. 1987. "An Algebraic Synthesis of the Foundations of Logic and Probability." *Information Sciences 42*, 187-237.

Dubois, D. and H. Prade. 1988. *Possibility Theory: a Approach to the Computerized Processing of Uncertainty*, Plenum Press, New York.

Goodman, I.R., H.T. Nguyen; and E.A. Walker. 1991. *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-free Conditioning*, North-Holland, New York.

Lipski, W. Jr. 1981. "On Databases with Incomplete Information." *Journal of the ACM, 28*, 41-70.

Moore, R.E. 1966. *Interval Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey.

Pawlak, Z. 1981. "Information Systems – Theoretical Foundations." *Information Systems 6*, 205-218.

Pawlak, Z. 1982. "Rough Sets." *International Journal of Computer and Information Sciences 11*, 341-356.

Roberts, F.S. 1979. *Measurement Theory*, Addison-Wesley, Reading, Massachusetts.

Ruspini, E.H. 1991. "Approximate Reasoning: Past, Present, Future." *Information Sciences 57-58*, 297-313.

Shafer, G. 1976. *A Mathematical Theory of Evidence. Princeton University Press, Princeton.*

Vakarelov, D. 1991. "A Modal Logic for Similarity Relations in Pawlak Knowledge Representation Systems." *Fundamenta Informaticae XV*, 61-79.

Wellman, M. and R.G. Simmons. 1988. "Mechanisms for Reasoning about Sets." In *Proceedings of the 7th AAAI National Conference on Artificial intelligence*, 398-402.

Yao, Y.Y. 1993. "Interval-set Algebra for Qualitative Knowledge Representation." In *Proceedings of the Fifth International Conference on Computing and Information*, 370-374.