DEMONSTRATIO MATHEMATICA
Vol. XXVII No 3–4 1994

Janusz A. Pomykała

ON SIMILARITY BASED APPROXIMATION OF INFORMATION

Dedicated to Professor Tadeusz Traczyk

There exists two basic intuitions related to our considerations – one corresponds to the notion of similarity and the other to the notion of approximation. Intuitively speaking, if the object x is similar to the object y then also the object y is similar to x. It is also reasonable to assume that every object x is similar to itself. Therefore, to the intuitive notion of similarity corresponds on the formal level the reflexive and symmetric relation (usually called tolerance).

As regards the approximation, assume that we have a subset X of the fixed set U. Sometimes we define in X the subset in some sense regular or closed (definable) with respect to given properties. This subset is often called the lower approximation of X. More generally, the method by which we construct the lower approximation for subsets of U (and dually – the upper approximation) is called the approximation operation. In particular such operations were considered in such formal structures as – relation systems, Boolean algebras, lattices, information systems, data bases etc.

This note is based on the notion of an information system (U, Ω, V, F) according to Pawlak. More precisely, we shall consider the nondeterministic system (U, Ω, V, F). Every set of attributes $P \subseteq \Omega$ determines a closure operator Ind^P on the one hand and a similarity approximation operation sim^P on the other. The class of Boolean algebras with added operators Ind^P determined by all sets of attributes was axiomatized by S. D. Comer in [1]. Our aim here is to obtain the axiomatization of the class of all Boolean algebras with added operations sim^P determined by all sets of attributes.

Now we recall basic notions, conventions and definitions. $P(U)$ denotes the family of all subsets of U. $P_n(U)$ is the family of all subsets of cardinality n of U. If $R \subseteq U \times U$ and $X \subseteq U$ then $RX = \{y \in U : \exists x \in X \; xRy\}$.

$B = (B, \cup, \cap, \sim, 0, 1)$ will be always a complete and atomic Boolean algebra. $At\, B$ will denote the set of all atoms of B. By h we denote the Stone's isomorphism (see Sikorski [5] p. 28).

An information system (deterministic) is a quadruple (U, Ω, V, f) where U is a set of objects, Ω stands for a set of attributes, V is a set of attributes values and $f : U \times \Omega \to V$ is a function (called information function), (see [2]). For every subset $P \subseteq \Omega$ an indiscernibility relation $\mathrm{Ind}(P) \subseteq U^2$ is defined in the following way : for any $x, y \in U$ $x \,\mathrm{Ind}(P) y$ iff $f(x,a) = f(y,a)$ for every $a \in P$ (see [2]).

A nondeterministic information system is an ordered tuple $S = (U, \Omega, V, F)$ where U is a set, Ω is a finite set, V is a set and F is a function $F : U \times \Omega \to P(V) \setminus \{o\}$.

For every $P \subseteq \Omega$ we define the following binary relation $\mathrm{sim}(P)$, called similarity relation : for $x, y \in U$ $x \,\mathrm{sim}(P) y$ iff $\forall a \in P F(x,a) \cap F(y,a) \neq \emptyset$ (see [3]).

CONVENTION. if $P = \{a\}$ i.e. P is a one element set, we shall write $\mathrm{sim}(a)$ instead of $\mathrm{sim}(\{a\})$.

For every set $X \subseteq U$ we define upper approximation of X in the following way

$$\mathrm{sim}^P X = \{u \in U : \exists x \in X, \quad u \,\mathrm{sim}(P) x\}$$

and the lower approximation of X

$$\mathrm{sim}_P X = -\mathrm{sim}^P(-X).$$

In other words sim_P is a dual operator to sim^P.

A structure $\mathbf{B} = \langle P(U), \cup, \cap, \sim, 0, U, \{\mathrm{sim}^P : P \subseteq \Omega\}\rangle$ (or $\langle \mathbf{P}(U), \mathrm{sim}^P \rangle_{P \subseteq \Omega}$ for short) is called the similarity based algebra of the Ω type derived from the information system S.

The next definition presents axioms for an abstract similarity approximation algebra (SAA) of type Ω. The idea is to abstract the properties of the operator sim^P as an operator T_P.

DEFINITION. Let us assume that the set Ω is finite. The structure $\mathbf{B} = \langle B, Tp \rangle_{P \subseteq \Omega}$ is a similarity (based) approximation algebra of type Ω if $T_P \in B^B$ for each $P \subseteq \Omega$ and the following conditions hold for all $x, y \in B$ and $P, Q \subseteq \Omega$:

0 B is complete atomic Boolean algebra,

1 $x \leq T_P x, \qquad 0 = T_P 0.$

2 $\forall x, y \in At\ B \qquad (y \leq T_P x \quad \text{iff} \quad x \leq T_P y).$

3 $\forall x \in B\ T_P(x) = \sup\{T_P(a) : a \leq x, \quad a \in At\ B\}.$

4 $\forall x \in At\ B \qquad T_{P \cup Q} x = T_P x \cap T_Q x.$

PROPOSITION 1. *The similarity algebra* $< P(U), \{\mathrm{sim}^P : P \subseteq \Omega\} >$ *satisfies conditions* 0–4.

THEOREM 1. *If* $< B, T_P >_{P \subseteq \Omega}$ *is a similarity approximation algebra then there exists an information system* (U, Ω, V, F) *such as*

$$T_P = \mathrm{sim}^P.$$

Proof. We define the system (U, Ω, V, F) in the following way $U = At\ B$; Ω is the same set as above i.e. it is the type of the similarity approximation algebra $< \mathrm{B}, T_P >_{P \subseteq \Omega}$, $V = P_2(U) \cup P_1(U)$ the family of subsets of U of cardinality at most 2. The function F will be defined in the following way, for $x \in U$ and $a \in \Omega$, $F : U \times \Omega \to P(V)$, $F(x, a) = \{\{x, y\} : x \leq T_a y\}$. We show now that for every $x, y \in U$, $P \subset \Omega$, $x \,\mathrm{sim}(P) y$ iff $x \leq T_P y$. Assume first that $x = y$. Then obviously $F(x, a) \cap F(y, a) = F(x, a) \neq \emptyset$, so $x \,\mathrm{sim}(a) x$ holds. On the other hand $x \leq T_a x$ in view of the axiom 1.

Next assume that $x \neq y$. We have:

$$F(x, a) = \{\{x, x'\} : x \leq T_a x'\}, \qquad F(y, a) = \{\{y, y'\} : y \leq T_a y'\}$$

therefore $F(x, a) \cap F(y, a) \neq \emptyset$ iff $\exists z_1, z_2 \{z_1, z_2\} = \{x, x'\}$ and $\{z_1, z_2\} = \{y, y'\}$

iff (i) $(z_1 = x = y'$ and $z_2 = y = x')$ or

(ii) $(z = x' = y$ and $z = x = y')$.

Assume that (i) holds. In case (ii) is satisfied the proof is analogical. Thus $\{z_1, z_2\} = \{x, y\} = \{y', x'\} \in F(x, a) \cap F(y, a)$ iff $x \leq T_a y$ and $y \leq T_a x$. We have proved that: $x \,\mathrm{sim}(a) y$ iff $x \leq T_a y$ and $y \leq T_a x$. Now for any $P \subseteq \Omega$ we have $x \,\mathrm{sim}(P) y$ iff $\forall a \in P\ F(x, a) \cap F(y, a) \neq \emptyset$ iff $\forall a \in P\ x \leq T_a y$ and $y \leq T_a x$. In view of the axiom 4 we obtain

$$x \leq \bigcap_{a \in P} T_a y = T_P y \quad \text{and} \quad \leq \bigcap_{a \in P} T_a x = T_P x.$$

So finally

$$x \,\mathrm{sim}(P) y \quad \text{iff} \quad x \leq T_P y \quad \text{and} \quad y \leq T_P x$$

or equivalently (see axiom 4) $x \,\mathrm{sim}(P) y$ iff $x \leq T_P y$. Now we recall the corollary from the Stone's representation theorem for Boolean algebras (see Sikorski [5] p. 28):

For every element x of a Boolean algebra B let $h(x)$ be the set of all atoms a of B such that $a \leq x$. The mapping h is a homomorphism of B into the field of all subsets of the set $X = h(1)$ of all atoms. If B is atomic, then h is an isomorphism.

Using this theorem we prove now the following fact: for every $x \in B$ $\mathrm{sim}^P(h(x)) = h(T_P(x))$. (Therefore, with respect to this isomorphism h we can write $\mathrm{sim}^P = T_P$).

$$\begin{aligned}\mathrm{sim}^P(h(x)) &= \mathrm{sim}^P(\{u \in U : u \leq x\}) = \\ &= \{z \in U : z\mathrm{sim}(P)u, \quad \text{for some} \quad u \leq x\} = \\ &= \{z \in U : \exists u \leq x,\ u \in U,\ z \leq T_P u\} = \\ &= \{z \in U : \exists u \leq x,\ u \in U,\ \forall a \in P \quad z \leq T_a u\}.\end{aligned}$$

On the other hand we have the following

$$\begin{aligned}h(T_P(X)) &= h(\sup\{T_P u : u \leq x \quad \text{and} \quad u \in U\}) = \\ &= h(\sup\{z : z \leq T_P u,\ u \leq x,\ u, z \in U\}) = \\ &= \{z : \exists u \leq x,\ u \in U,\ z \leq T_P u,\ z \in U\} = \\ &= \{z \in U : \exists u \leq x,\ u \in U(\forall a \in P,\ z \leq T_a u)\}. \blacksquare\end{aligned}$$

The system constructed in the above theorem will be called the information system determined by the algebra $< B, T_P >_{P \subseteq \Omega}$.

COROLLARY. *If $\langle U, \Omega, V, F\rangle$ is the system determined by the algebra $\langle B, T_P\rangle_{P\subseteq\Omega}$ and $\langle P(U), \mathrm{sim}^P\rangle_{P\subseteq\Omega}$ is the algebra derived from the system $\langle U, \Omega, V, F\rangle$ then $\langle B, T_P\rangle_{P\subseteq\Omega}$ is isomorphic to $\langle P(U), \mathrm{sim}^P\rangle_{P\subseteq\Omega}$.*

Now we will prove that if $\langle P(U), \mathrm{sim}^P\rangle$ is derived from $\langle U, \Omega, V, F\rangle$ and if $\langle U', \Omega', V', F'\rangle$ is the system determined by the algebra $\langle B, T_P\rangle_{P\subseteq\Omega}$then the systems $\langle U, \Omega, V, F\rangle$ and $\langle U', \Omega', V,' F'\rangle$ are isomorphic. Of course first we have to introduce the definition of isomorphism between information systems.

DEFINITION. The system $< U, \Omega, V, F >$ is isomorphic to the system $\langle U', \Omega', V', F'\rangle$ iff the following conditions are satisfied:

(a) there exists the mapping

$$i_U : U \xrightarrow[\text{onto}]{1-1} U',$$

(b) there exists the mapping

$$i_\Omega : \Omega \xrightarrow[\text{onto}]{1-1} \Omega',$$

(c) for every $a \in \Omega$ and for every $x, y \in U$,

$$(F(x,a) \cap F(y,a) \neq \emptyset \quad \text{iff} \quad F'(i_U(x),\ i_\Omega(a)) \cap F'(i_U(y),\ i_\Omega(a)) \neq \emptyset).$$

Let us mention that the relation of isomorphism between the information systems is the equivalence relation.

THEOREM 2. *If $\langle P(U), \mathrm{sim}^P\rangle_{P\subseteq\Omega}$ is the similarity approximation algebra derived from the system $\langle U, \Omega, V, F\rangle$ and if $\langle U', \Omega', V', F'\rangle$ is the information* *system determined by the algebra* $\langle B, T_P\rangle_{P\subseteq\Omega}$ *then the systems* $\langle U,\Omega,V,F\rangle$ *and* $\langle U',\Omega',V',F'\rangle$ *are isomorphic.*

Proof. The mappings i_U, i_Ω are defined as follows: $i_U : U \to U'$, $i_U(x) = \{x\}$; $i_\Omega(a) = a$ for every $a \in \Omega$. As a consequence the conditions (a) and (b) are satisfied. Now, let us assume that for $a \in \Omega$ and $x, y \in U$ we have $F(x,a) \cap F(y,a) \neq \emptyset$. In view of the definition of the similarity approximation algebra determined by the information system we infer that: $U' = P_1(U)$, $\Omega' = \Omega$, and $F'(\{x\},a) = \{\{\{x\},\{y\}\} : \{x\} \leq \mathrm{sim}^a\{y\}\}$. This implies that $F'(\{x\},a) \cap F'(\{y\},a) \neq \emptyset$, in other words $F'(i_U(x), i_\Omega(a)) \cap F'(i_U(y), i_\Omega(a)) \neq \emptyset$. ∎

Finally, let us observe that every tolerance relation may be represented as a similarity relation in a properly chosen information system. Namely, it holds:

PROPOSITION 2. *If* $T \subseteq U \times U$ *is the tolerance relation and the information system* (U,Ω,V,F) *is defined in the following way*: $\Omega = \{a\}$, $V = U$, $F(x,a) = \{\{x,y\} : xTy\}$, *then*: $\mathrm{sim}(a) = T$. ∎

This note is a part of the presentation given during Banach semester, 1991.

Acknowledgement. This work is inspired by the stimulating discussion with Professor S. D. Comer. The author is also grateful to Professor T.B. Iwiński.

Appendix

Now we have observed that every tolerance relation (i.e. reflexive and symmetric relation) may be represented as a similarity relation in a properly chosen information sysyem. Therefore we may equivalently formulate our results for arbitrary family $\{\tau a : a \in \Omega\}$ of tolerances in U.

In what follows we formulate some properties of the approximation operators based on tolerances. We recall that our basic intuitions are the following:

(i) an approximation operation may be used to classify sets of informations, data or sets of objects;

(ii) these operations give us the possibility to approximate a knowledge or to decide the membership question with respect to tolerance (or equivalence) relations.

Let us consider a tolerance $\tau : U \times U$ and define the operation $\tau : P(U) \to P(U)$ in the following standard way: for every $X \subset U$

$$\tau X = \{y \in U : \exists x \in X \; x\tau y\}.$$

PROPOSITION 1.

(a) $\tau\emptyset = \emptyset$, $\tau U = U$

(b) $X \subseteq \tau X$

(c) $X \subseteq Y$ *implies* $\tau X \subseteq \tau Y$

(d) $\tau(X \cup Y) = \tau X \cup \tau Y$

(e) $\tau(X \cap Y) \subseteq \tau X \cap \tau Y$

(f) $\tau(-X) \supseteq -\tau(X)$

(g) $\tau X \cap \tau(-X) = \emptyset$ *iff* $\tau X = X$ *iff* $\tau(-X) = -X$

(h) $\forall X\ \ \tau X \cap \tau(-X) = \emptyset$ *iff* $\forall X\ \ \tau X = X$ *iff* $\tau = \mathrm{Id}$ *iff* $\forall x\ \ \tau(\{x\}) \cap \tau(-\{x\}) = \emptyset$

(i) $\tau\tau X \supseteq \tau X$.

The operation δ conjugated to τ is defined in the following way: for any $X \subseteq U$

$$\delta X = -\tau(-X).$$

PROPOSITION 2.

(a) $\delta\emptyset = \emptyset$, $\delta U = U$

(b) $X \supseteq \delta X$

(c) $X \subseteq Y$ *implies* $\delta X \subseteq \delta Y$

(d) $\delta(X \cup Y) \supseteq \delta X \cup \delta Y$

(e) $\delta(X \cap Y) \subseteq \delta X \cap \delta Y$

(f) $\delta(-X) \supseteq -\delta(X)$

(g) $\delta X \cap \delta(-X) = \emptyset$ *iff* $\delta X = X$ *iff* $\delta(-X) = -X$

(h) $\forall X\ \ \delta X \cap \delta(-X) = \emptyset$ *iff* $\forall X\ \ \delta X = X$ *iff* $\delta = \mathrm{Id}$ *iff* $\forall x\ \ \delta(\{x\}) \cap \delta(-\{x\}) = \emptyset$

(i) $\delta\delta X \subseteq \delta X$

(j) $\tau\delta X \subseteq X \subseteq \delta\tau X$

(k) $\tau_1 \subseteq \tau_2$ *implies* $\tau_1 X \subseteq \tau_2 X$

(l) $\tau_1 X \subseteq \tau_2 X$ *implies* $\delta_2 X \subseteq \delta_1 X$.

Remark. Even if $X \cap Y = \emptyset$ and if δ is an equivalence it is not true in general that $\delta X \cup \delta Y = \delta(X \cup Y)$.

PROPOSITION 3.

(a) $\tau X \cup \delta(-X) = U$

(a) $\tau X \cap \delta(-X) = \emptyset$

(c) $-\tau X = \delta(-X)$

(d) $-\delta X = \tau(-X)$

(e) *if* $\tau_3 = \tau_1 \cap \tau_2$ *then* $\tau_3 X \subseteq \tau_1 X \cap \tau_2 X$.

Remark. We can interpret (e) in the following way: the intersection of the above operators does not correspond strictly to the intersection of the corresponding relations.

Now let us assume that U, Ω are nonempty sets, $A, B, C, \ldots \subseteq \Omega$, $a, b, c, \ldots \in \Omega$, $X, Y, Z, \ldots \subseteq U$, $x, y, z, \ldots \in U$. Let us assume also that for every $a \in \Omega$ the relation $\tau_a \subseteq U \times U$ is reflexive and symmetric. We define the relation τ_A for every set $A \subseteq \Omega$ in the following way

$$\tau_A = \bigcap_{a \in A} \tau_a.$$

Let us observe that τ_A is also a tolerance relation.

At the moment we examine the behavior of the relations and the corresponding operations τ_A and τ_a, for $a \in A$, with respect to the boolean operations in $P(U)$.

Remark. There can be defined two more approximation operations, namely

$$\tau_A^1 X = \bigcap_{a \in A} \tau_a X$$

and

$$\tau_A^2 X = \bigcup_{a \in A} \tau_a X.$$

We shall not consider them in the sequel. Let us only state that in general $\tau_A x$ is not equal to $\tau_A^1 x$.

PROPOSITION 4.
(a) $\forall_{a \in A}\ \tau_A \subseteq \tau_a$
(b) $\tau_A X \subseteq \bigcap_{a \in A} \tau_a X$
(c) $\forall_{a \in A}\ \tau_A X \subseteq \tau_a X$
(d) *if* $X = \{x\}$ *then* $\tau_A X = \bigcap_{a \in A} \tau_a X$
(e) $\tau_A(X \cup Y) \subseteq \bigcap_{a \in A} \tau_a X \cup \bigcap_{a \in A} \tau_a Y$
(f) $\tau_A(X \cup Y) = \tau_a X \cup \tau_a Y$
(g) $\tau_A(X \cap Y) \subseteq \tau_a X \cap \tau_a Y$
(h) $\forall_{a,b \in A} \tau_A(X \cap Y) \subseteq \tau_a X \cap \tau_b Y$
(i) $\tau_A(X \cap Y) \subseteq \bigcap_{a \in A}(\tau_a X \cap \tau_a Y)$
(j) $\tau_A(-X) \supseteq -X \supseteq -\tau_A X$.

Now we investigate the properties of the operators $\tau_A, \tau_B, \ldots$ with respect to the boolean operations in $P(\Omega)$.

PROPOSITION 5.
(a) $\tau_{A \cup B} = \bigcap_{a \in A} \tau_a \cap \bigcap_{b \in B} \tau_b$
(b) $\tau_{A \cup B} X \subseteq \tau_A \cap \tau_B$

(c) $\tau_{A\cup B} \subseteq \bigcap_{a\in A\cup B} \tau_a X = \bigcap_{a\in B} \tau_a X$
(d) $A \subseteq B$ *implies* $\tau_A \supseteq \tau_B$
(e) $\tau_{A\cap B} \supseteq \tau_A \cap \tau_B$
(f) $\tau_{-A\cup A} = \tau_{\Omega}$.

There are some properties of our operators connecting the boolean structures of $P(U)$ and $P(\Omega)$.

PROPOSITION 6.
(a) $\tau_{a\cup B}(X \cup Y) = \tau_{A\cup B}X \cup \tau_{A\cup B}Y \subseteq \tau_A X \cap \tau_B X \cup \tau_A Y \cap \tau_B Y$
(b) $\tau_{a\cap B}(X \cup Y) = \tau_{A\cap B}X \cup \tau_{A\cap B}Y \supseteq \tau_A X \cup \tau_B X \cup \tau_A Y \cup \tau_B Y = \tau_A(X \cup Y) \cup \tau_B(X \cup Y)$
(c) $\tau_{a\cup B}(X \cap Y) = \tau_{A\cup B}X \cap \tau_{A\cup B}Y \subseteq \tau_A X \cap \tau_B X \cap \tau_A Y \cap \tau_B Y$
(d) $\tau_{-A}(X \cup Y) = \tau_{-A}X \cup \tau_{-A}Y$
(e) $\tau_{-A}(X \cap Y) \subseteq \tau_{-A}X \cap \tau_{-A}Y$
(f) $\tau_{-A}(-X) \supseteq -\tau_{-A}X$.

PROPOSITION 7.
(a) $\tau_{\emptyset} = U \times U$
(b) $\tau_{\emptyset}X = U$ *for* $X \neq \emptyset$
(c) $\tau_{\emptyset} = \emptyset$
(d) $\tau_A\emptyset = \emptyset$
(e) $\tau_A U = U$
(f) $\delta_{\emptyset}U = U$
(g) $\delta_{\emptyset}X = \emptyset$ *for* $X \neq U$
(h) $\delta_A U = U$
(i) $\delta_A\emptyset = \emptyset$.

It is now a standard matter to formulate a modal logic with a family of possibility and necessity operators, which can be interpreted in the similarity based algebra (of the corresponding type) or — equivalently — in the tolerance based algebra.

Let us finally notice that every relational system can be represented in terms of the information system.

References

[1] S. D. Comer, *An algebraic approach to the approximation of information*, Fund. Informat. 14 (1991), 492–503.

[2] Z. Pawlak, *Information system, theoretical fundations*, Information Syst. 6 (1981), 205–218.

[3] E. Orłowska, *Logic of nondeterministic information*, Studia Logica (Wrocław), 44 (1985).

[4] J. A. Pomykała, *Some remarks on approximation*, Demonstratio Math. 24 (1991), 95–104.
[5] R. Sikorski, *Boolean Algebras*, Springer Verlag, 1964.
[6] B. Jonsson, A. Tarski, *Boolean Algebras with Operators*, part I. Transactions of the AMS, 1951.
[7] T. Traczyk, *Common extension of Boolean informational systems*, Fund. Informat. 2 (1978), 63–70.

MILITARY UNIVERSITY OF TECHNOLOGY
Kaliskiego 2
01-489 WARSZAWA 49, POLAND

Received August 9, 1993.