BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
MATHEMATICS
Vol. 32, No. 9–10, 1984

THEORETICAL COMPUTER SCIENCE

Measurement and Indiscernibility

by

Ewa ORŁOWSKA and Zdzisław PAWLAK

Presented by Z. PAWLAK on January 24, 1984

Summary. In the paper we discuss some aspects of imperfectness of measurement. We present a relationship between the basic notions of the theory of measurement and the theory of rough sets.

1. Introduction. Our object in this paper is to present one general approach to the interpretation of the concepts of measurement theory, an approach based on the theory of rough sets (Pawlak [4]). We draw attention to the indiscernibility determined by a measurement function and we discuss some important aspects of imperfectness of measurement. Such an approach gives indication of how one is able to reconstruct or interpret empirical data and their properties on the base of results of measurement. The resulting theory provides the grounds for data analysis.

2. Approximation space. In the present section we bring the basic notions introduced in Konrad, Orłowska and Pawlak [2], Orłowska [3], Pawlak [4].

By an approximation space we mean a system

$$S = (U, I)$$

where U is a non-empty set called universe of discourse,

I is an equivalence relation on set U called indiscernibility relation.

An indiscernibility relation is considered to be a formal counterpart of a perception or observation ability. According to the set-theoretical view we shall identify properties of elements of universe U with its subsets. Given a subset X of U, we say that

X is definable (finitely definable) in space S iff X is the empty set or X can be represented as a union of some (finite number) of the equivalence classes of relation I.

Usually it is a case that not all the subsets of U are definable in S. Given a non-definable subset X of U, we are not able to pick up a sharp boundary between X and its complement $U-X$, we perceive X with some tolerance. The limits of this tolerance are determined by a pair $\overline{I}X$ and $\underline{I}X$ of definable sets called upper and lower approximation of X, respectively. The formal definition of these sets is as follows:

$\underline{I}X = \{x \in U\text{:for all } y \in U \text{ if } (x, y) \in I \text{ then } y \in X\}$

$\overline{I}X = \{x \in U\text{: there is a } y \in U \text{ such that } (x, y) \in I \text{ and } y \in x\}$

Clearly, if x is definable in S then $\underline{I}X = \overline{I}X = X$.

In the following we list some properties of approximations.

Fact 2.1.

(a) $\underline{I}(X \cap Y) = \underline{I}X \cap \underline{I}Y$,

(b) $\underline{I}X \subseteq X$,

(c) $\underline{I}\underline{I}X = \underline{I}X$,

(d) $\underline{I}U = U$.

Fact 2.2.

(a) $\overline{I}(X \cup Y) = \overline{I}X \cup \overline{I}Y$,

(b) $X \subseteq \overline{I}X$,

(c) $\overline{I}\overline{I}X = \overline{I}X$,

(d) $\overline{I}\emptyset = \emptyset$.

It follows that the algebra $P(U)$ of all the subsets of U with the additional operators $\underline{I}$ and $\overline{I}$ is a topological field of sets, where $\underline{I}$ is the interior operation and $\overline{I}$ is the closure operation.

Fact 2.3.

(a) $\underline{I}X = -\overline{I}(-X)$,

(b) $\overline{I}X = -\underline{I}(-X)$,

(c) if $X \subseteq Y$ then $\underline{I}X \subseteq \underline{I}Y$ and $\overline{I}X \subseteq \overline{I}Y$.

An approximation space (U, I) is said to be selective iff I is an identity on set U.

Fact 2.4. The following conditions are equivalent:

(a) A space $S = (U, I)$ is selective.

(b) For any $X \subseteq U$ $\underline{I}X = X = \overline{I}X$.

(c) For any $X \subseteq U$ X is definable in space S.

Treating a subset X of U as a property of elements of U we define sets of positive, negative and borderline instances of this property as follows:

an element $x \in U$ is a positive instance of a property represented by X iff $x \in \underline{I}X$,

an element $x \in U$ is a negative instance of a property represented by X iff $x \in U - \overline{I}X$,

an element $x \in U$ is a borderline instance of a property represented by X iff $x \in \overline{I}X - \underline{I}X$.

Hence set $\underline{I}X$ consists of the elements of the universe U which definitely, up to the equivalence I, obey a property corresponding to X. Set $U-\bar{I}X$ consists of the elements which definitely, up to the equivalence I, do not obey this property. The borderline region $\bar{I}X-\underline{I}X$ is a doubtful area. It consists of the objects for which perception ability determined by I is not sufficient to decide whether they obey the given property.

The indiscernibility relation I induces the equivalence relations in the set $P(U)$ which are considered to be approximate equalities of sets:

$$X \underline{eq} Y \text{ iff } \underline{I}X = \underline{I}Y,$$
$$X \overline{eq} Y \text{ iff } \bar{I}X = \bar{I}Y,$$
$$X \underline{\overline{eq}} Y \text{ iff } X \underline{eq} Y \text{ and } X \overline{eq} Y.$$

Given a subset X of U which is non-definable in space S, our perception restricted by I causes X to be perceived as a family of those sets which cannot be distinguished from X by means of relation *eq*. Hence a vague property determines not a single set of elements falling under the property but a family of sets which can be identified with this property up to indiscernibility I.

In a natural way we can extend the given notions to n-ary relations which can be treated as properties of n-tuples of elements of a universe (Pawlak [5.6]).

3. Measurement. In the measurement theory (Scott and Suppes [8]) empirical data which are to be measured are identified with a relational system

$$DS = (U, \mathrm{Rel}(U), \mathrm{Fun}(U))$$

where U is a non-empty set called a universe of data items,

Rel (U) is a family of relations on set U

Fun (U) is a family of functions on set U.

Measurement is considered to be a homomorphism f from DS into a numerical structure RS similar to DS:

$$f: U \to Q$$
$$RS = (Q, \mathrm{Rel}(Q), \mathrm{Fun}(Q))$$

for any n-ary relation $r \in \mathrm{Rel}(U)$, $n \geqslant 1$, there is an n-ary relation $r' \in \mathrm{Rel}(Q)$ such that for all $x_1, \ldots, x_n \in U$

$$(x_1, \ldots, x_n) \in r \text{ iff } (f(x_1), \ldots, f(x_n)) \in r'$$

for any n-argument function $o \in \mathrm{Fun}(U)$, $n \geqslant 0$, there is an n-argument function $o' \in \mathrm{Fun}(Q)$ such that for all $x_1, \ldots, x_n \in U$

$$o(x_1, \ldots, x_n) = o'(f(x_1), \ldots, f(x_n)).$$

Our approach differs from that developed by Scott and Suppes. First, we do not assume that measurement is a homomorphism from the empirical structure into a result structure, and second we do not assume that the results structure is a numerical one. Observe, that if f was a homomorphism then the fact that measurement could not distinguish all the objects from U would not influence the relationships between objects expressed by means of relations. Consider the following example.

Let $U = \{x_1, x_2, x_3, x_4\}$ and let $\mathrm{Rel}(U)$ contain the one-place relation $r_1 = \{x_1, x_2, x_3\}$ and the binary relation $r_2 = \{(x_1, x_2), (x_2, x_3)\}$. Let $f: U \to Q = \{q_1, q_2, q_3\}$ be defined as follows: $f(x_1) = f(x_2) = q_1$, $f(x_3) = q_2$, $f(x_4) = q_3$. Then for any one-place relation $p_1 \subseteq Q$ and any binary relation $p_2 \subseteq QxQ$ the conditions $x \in r_1$ iff $f(x) \in p_1$ and $(x, y) \in r_2$ iff $(f(x), f(y)) \in p_2$ are not satisfied.

For the sake of simplicity we confine ourselves to structures with one-place relations. To extend the presented approach to arbitrary relations and functions is only a technical matter. The respective results can be obtained on the base of the notions of rough relation and rough function introduced in Pawlak [5, 6].

Let data structure DS be given:

$$DS = (U, R)$$

where $R \subseteq P(U)$.

By a measurement of structure DS we man a function $f: U \to Q$, where Q is a non-empty set. For $X \subseteq U$ let $f(x)$ denote the image of set X determined by f:

$$f(X) = \{q \in Q: \text{there is an } x \in X \text{ such that } f(x) = q\}.$$

Then by the result structure we mean the system

$$RS = (Q, \{f(X): X \in R\}).$$

In the process of analysing empirical data on the base of measurement results we have to take into account the imperfectness of measurement. For the purpose we define the approximation space determined by a measurement. Let I_f be the equivalence relation on set U defined as follows:

$$(x, y) \in I_f \text{ iff } f(x) = f(y).$$

Let us consider the approximation space

$$S_f = (U, I_f).$$

In this approximation space empirical data can be identified up to the indiscernibility relation I_f. This fact influences our perception of relationships between these data. It seems natural to use the concepts presented in section **2** to express consequences of inexactness of measurement. We say that

a set $X \subseteq U$ is f-measurable iff X is definable in space S_f.

For data sets which are not measurable we can define their approximations in space S_f. Consider the following example. Let data structure DS consist of set R^+ of all non-negative real numbers and the family $\{\langle 0; r)\}_{r\in R^+}$. Let measurement $f: R^+ \to \omega$ be defined as follows:

$$f(x) = \text{entier of } X.$$

Then the result structure consists of set ω of natural numbers and the family of all the finite subsets of ω. Equivalence classes of relation I_f are listed below

$$\langle 0; 1), \langle 1; 2), \langle 2; 3), \ldots$$

We have

$$\underline{I}_f \langle 0; r) = \langle 0; f(r))$$
$$\overline{I}_f \langle 0; r) = \langle 0; f(r)+1)$$

It follows that the only f-measurable sets in space S_f are intervals of the form $\langle 0; n)$ for $n \in \omega$.

The following conditions reflect the influence of inexactness of measurement on the results of masurement. For $P \subseteq Q$ let $f^{-1}(P)$ denote the set of those elements of U whose images belong to P:

$$f^{-1}(P) = \{x \in U : f(x) \in P\}$$

Fact 3.1.

(a) For any $q \in Q$ set $f^{-1}\{q\}$ is an equivalence class of relation I_f.
(b) $X \subseteq f^{-1}(f(X))$ for any $X \subseteq U$.
(c) If X is f-measurable then $X = f^{-1}(f(X))$.
(d) $f^{-1}(f(X)) = \overline{I}_f X$.

Fact 3.2.

(a) $f(\underline{I}_f X) \subseteq f(X)$.
(b) $f(X) = f(\overline{I}_f X)$.
(c) If $X eq_f Y$ then $f(X) = f(Y)$.

These facts show that in general it is not possible to reconstruct data from the results of measurement. Data items can be identified up to indiscernibility I_f and properties (subsets) of entities from U can be recognized with the tolerance determined by their lower and upper approximations.

A measurement f is said to be perfect iff function f is injection.

Fact 3.3. The following conditions are equivalent:

a) Measurement f is perfect.
b) Approximation space S_f is selective.
c) For any set $X \subseteq U$ $f^{-1}(f(X)) = X$.

Consider the following example

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

Family R consists of the following subsets of U:

$$X_1 = \{x_1, x_3, x_4\}$$
$$X_2 = \{x_1, x_3, x_6, x_7, x_8\}$$
$$X_3 = \{x_4, x_5\}$$
$$Q = \{\text{red, blue, green}\}$$

Measurement f determines colours of objects from U:

$$f(x_1) = f(x_3) = f(x_7) = \text{red},$$
$$f(x_2) = f(x_6) = f(x_8) = \text{blue},$$
$$f(x_4) = f(x_5) = \text{green}.$$

It follows that sets X_1 and X_2 are not f-measurable and set X_3 is f-measurable.

We have:

$$f(X_1) = \{\text{red, green}\}$$
$$\underline{I}_f X_1 = \emptyset \quad \bar{I}_f X_1 = \{x_1, x_3, x_4, x_5, x_7\}$$
$$f^{-1}(\{\text{red, green}\}) = \bar{I}_f X_1$$
$$f(X_2) = \{\text{red, blue}\}$$
$$\underline{I}_f X_2 = \{x_1, x_3, x_7\} \quad \bar{I}_f X_2 = \{x_1, x_2, x_3, x_6, x_7, x_8\}$$
$$f^{-1}(\{\text{red, blue}\}) = \bar{I}_f X_2$$
$$f(X_3) = \{\text{red}\}$$
$$\underline{I}_f X_3 = X_3 = \bar{I}_f X_3$$
$$f^{-1}(\{\text{red}\}) = X_3.$$

Approximation space S_f reflects our view of data determined by the results of measurement F. Since measurement f is not necessarily perfect, that is some data items cannot be distinguished by I_f then, in general, data have not uniquely determined counterparts in space S_f.

In the following we introduce functions $\underline{m}_f$ and $\bar{m}_f$ which assign non-negative integers to those subsets of universe U which have finitely definable approximations in space S_f. The quotient $n_f(X)$ of the values of these functions for a set X reflects degree of inexactness of the measurement of X by f in a sense.

Let Y/I_f denote the set of those equivalence classes of relations I_f which are included in Y.

$$\underline{m}_f(X) = \text{card } \underline{I}_f X/I_f,$$
$$\bar{m}_f(X) = \text{card } \bar{I}_f X/I_f,$$
$$n_f(X) = \frac{\underline{m}_f(X)}{\bar{m}_f(X)}.$$

For example, for the entier measurement function we have

$$\underline{m}_f\langle 0; r) = f(r)$$
$$\bar{m}_f\langle 0; r) = f(r) + 1.$$

4. Families of measurement. In this section we present counterparts of the notions of regular, admissible and homogeneous family of measurements investigated in the measurement theory (Bromek, Moszyńska and Prażmowski [1]) and we characterize these notions in terms of approximation spaces generated by measurements.

Given a family F of measurements from U to Q let $\mathrm{Ind}(F)$ be the family of all the indiscernibility relations determined by the functions from F.

A measurement f is said to be regular with respect to a family F iff for every measurement $g \in F$ we have $I_f \subseteq I_g$.

Fact 4.1. For any family F any injection $f \in F$ is regular with respect to F.

A family F is said to be regular iff each measurement $f \in F$ is regular with respect to F.

Fact 4.2. The following conditions are equivalent:

(a) Family F is regular.

(b) For any $f, g \in F$ $I_f = I_g$.

Hence a regular family of measurements consists of measurements which have the same measurability power.

Given a measurement $f \in F$, a function $h: f(U) \to Q$ is said to be admissible for f with respect to F iff $I_{hf} \in \mathrm{Ind}(F)$.

Fact 4.3. For any measurement $f: U \to Q$ and for any function $h: f(U) \to Q$ we have $I_f \subseteq I_{hf}$.

Let $A_F(f)$ denote the family of all the functions admissible for f with respect to F.

A family F is said to be homogeneous iff for any f and $g \in F$ we have $A_F(f) = A_F(g)$.

Let F be a family of measurements, we define the order in F as follows:

$$f \leqslant g \text{ iff } I_f \subseteq I_g.$$

Relation $\leqslant$ induces an equivalence E on family F:

$$(f, g) \in E \text{ iff } f \leqslant g \text{ and } g \leqslant f.$$

Equivalence classes of relation E are called types.

Fact 4.4. The following conditions are equivalent:

(a) A family F of measurements is regular.

(b) For any $f \in F$, F is the type generated by f.

Consider the following example.

$$U = \{x_1, x_2, x_3, x_4, x_5\},$$
$$Q = \{q_1, q_2, q_3\},$$
$$F = \{f_1, f_2, f_3\}, \quad G = \{f_2, f_3\}.$$
$$f_1(x_1) = f_1(x_5) = q_1,$$
$$f_1(x_2) = f_1(x_4) = q_2,$$
$$f_1(x_3) = q_3,$$

$$f_2(x_1) = f_2(x_3) = f_2(x_5) = q_1,$$
$$f_2(x_2) = f_2(x_4) = q_2,$$
$$f_3(x_1) = f_3(x_3) = f_3(x_5) = q_3,$$
$$f_3(x_2) = f_3(x_4) = q_2.$$

We have

f_1 is regular with respect to F,
G is regular.

Let $h: Q \to Q$ be defined as follows:

$$h(q_1) = h(q_3) = q_3,$$
$$h(q_2) = q_2.$$

We have $I_{hf_1} = I_{f_2} \in \mathrm{Ind}(F)$ and therefore h is admissible for f_1 with respect to F. Relation $\leqslant$ in family F consists of the following pairs:

$$(f_1, f_2)\ (f_1, f_3)\ (f_i, f_i) \text{ for } i = 1, 2, 3.$$

Relation E generates the following types in family F:

$$\{f_1\}\ \{f_2, f_3\}.$$

INSTITUTE OF COMPUTER SCIENCE, POLISH ACADEMY OF SCIENCES, PKIN, 00-901 WARSAW
(INSTYTUT PODSTAW INFORMATYKI, PAN)
COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF NORTH CAROLINA, CHARLOTTE, N.C. 28223, (USA)

REFERENCES

[1] T. Bromek, M. Moszyńska, K. Prażmowska, *Concerning basic notions of the measurement theory*, Czechoslov. Math. Journal, [to appear].

[2] E. Konrad, E. Orłowska, Z. Pawlak, *Knowledge representation systems*, ICS PAS Reports 433, 1981.

[3] E. Orłowska, *Semantics of vague concepts*, ICS PAS Reports 469, 1982.

[4] Z. Pawlak, *Rough sets*, International Journal of Information and Computer Sciences, **11** (1982), 341–356.

[5] Z. Pawlak, *Rough relations*, ICS PAS Reports 435, 1981.

[6] Z. Pawlak, *Rough functions*, ICS PAS Reports 467, 1981.

[7] F. S. Roberts, *Measurement theory with applications to decisionmaking, utility and the social sciences*, Addison Wesley Publ. Co., 1979.

[8] D. Scott, P. Suppes, *Foundational aspects of theories of measurement*, Journal of Symbolic Logic, **28** (1958), 113–128.

Э. Орловска, З. Павляк, **Измерение и неразличимость**

В настоящей работе рассматриваются некоторые элементы анализа данных на основе их измерения. Приводится интерпретация основных понятий теории измерения в теории приблизительных множеств.