About Tolerance and Similarity Relations in Information Systems

J.A. Pomykała

Manager Academy, Society for Trade Initiatives
ul. Kawęczyńska, Warsaw, Poland
pomykala_andrzej@mac.edu.pl

Abstract. The tolerance and similarity relations in information systems are considered. Some properties and connections are shown and the relation to the dependency of attributes in relational databases is developed. In particular a new definition of similarity dependency of attributes is formulated.

Keywords. Information system, dependency of attributes, similarity relation

1 Introduction

The notion of information systems formulated by Pawlak and developed by him and co-workers (see eg. Orłowska [10], Skowron [21], Dűntsch and Gediga [4], Vakarelov [24]) and Słowiński [22] is now a well developed branch of data analysis formalisms. It is strongly related (but different) to relational database theory on the one hand and to the fuzzy sets theory on the other.

In this paper we propose new notion of similarity of systems and we formulate some properties of attribute dependency, expressed in the language of weak and strong similarity relations $\mathrm{sim}\,(X)$ and $\mathrm{Sim}\,(X)$, respectively. This note can be seen as a step toward the solution of the open problem formulated in Orłowska [10]. Some results in this paper were presented in the proceedings of the Sixth International Workshop of Relational Methods in Computer Science, the Netherlands, October 2001.

2 Information Systems

We recall again that the aim of this paper is twofold: first give a broad motivation for studying similarity of systems and similarity relations (or equivalently tolerance relations) in geometry, and logic, and second give a new definition of dependency of attributes in RDB (relational database) and IST (information systems theory).

Any collection of data specified as a structure (O, A, V, f) such that O is a nonempty set of objects, A is a nonempty set of attributes, V is a nonempty

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 175–182, 2002.

set of values and f is a function of $O \times A$ into $2^{V \setminus \{\emptyset\}}$, is referred to as an information system.

In this paper we assume that with every attribute $a \in A$ is related a tolerance relation (i.e. reflexive and symmetric relation) $\tau(a)$. In most cases this relation shall be defined in the following way. Let $a \in A$ and $B \subseteq A$:

$$\begin{aligned}
&\mathrm{Sim}\,(a)xy \quad \text{iff} \quad f(x,a) \cap f(y,a) \neq \emptyset \\
&\mathrm{sim}\,(a)xy = \mathrm{Sim}\,(a)xy, \\
&\mathrm{Sim}\,(B)xy \quad \text{iff} \quad \forall b \in B \;\; \mathrm{Sim}\,(b)xy \\
&\mathrm{sim}\,(B)xy \quad \text{iff} \quad \exists b \in B \;\; \mathrm{sim}\,(b)xy.
\end{aligned}$$

$\mathrm{Sim}\,(B)$ is called (strong) similarity relation and $\mathrm{sim}\,(B)$ is called weak similarity with respect to the set of attributes $B \subseteq A$.

The set $\{f(x,a) : a \in A\}$ shall be called an information about the object x, in short a record of x or a row determined by x. We shall say that two records determined by x, y are strongly τ-similar iff $\forall a \in A \;\, f(x,a)\tau(a)f(y,a)$. We will also consider the case when the above notion is restricted to a set $B \subseteq A$ i.e. two records $\{f(x,a) : a \in B\}$ and $\{f(y,b) : b \in B\}$ are similar with respect to the set $B \subseteq A$ iff

$$\forall b \in B\, f(x,b)\tau(b)f(y,b).$$

We shall say that two objects (records) x, y are weakly τ-similar if for some attribute $a \in A$, values $f(x,a)$, $f(y,a)$ are similar with respect to $\tau(a)$.

In symbols: $\exists b \in B f(x,b)\tau(b)f(y,b)$. We denote strong relation by $\tau(B^{\wedge})$ and weak one by $\tau(B^{\vee})$, respectively.

We can express special kind of similarity of records by formulating the proper query in the system. It is however strongly determined by the possibilities of a given RDB system. By analogy to indiscernibility matrices (see Skowron, Rauszer [21]) we propose to use similarity matrices. We begin with the definition:

the set of attributes $Y \subseteq A$ depends on the set $X \subseteq A$ with respect to the similarity relation Sim *if and only if*

$$\mathrm{Sim}\,(X) \subseteq \mathrm{Sim}\,(Y).$$

We shall write in symbols

$$X \xrightarrow{S} Y \quad \text{or} \quad X \xrightarrow{\mathrm{Sim}} Y.$$

In the same way we can define dependency of attributes with respect to weak similarity relation sim :

$$X \xrightarrow{s} Y \quad \text{iff} \quad \mathrm{sim}\,(X) \leq \mathrm{sim}\,(Y)$$

(here $\leq$ is the usual inclusion relation between binary relations).

In other words, $X \xrightarrow{S} Y$ if strong similarity of objects with respect to the set of attributes X implies strong similarity of objects with respect to the set of attributes Y.