Information Sciences 442–443 (2018) 72–102

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Granular computing on information tables: Families of subsets and operators

G. Chiaselotti*, T. Gentile, F. Infusino

Department of Mathematics and Informatics, University of Calabria, Via Pietro Bucci, Cubo 30B, Arcavacata di Rende (CS) 87036, Italy

A R T I C L E I N F O

Article history:
Received 4 July 2017
Revised 3 February 2018
Accepted 18 February 2018
Available online 21 February 2018

Keywords:
Granular computing
Information tables
Set systems
Set operators
Closure systems
Dependency

A B S T R A C T

In this work we use the granular computing paradigm to study specific types of families of subsets, operators and families of ordered pairs of sets of attributes which are naturally induced by information tables. In an unifying perspective, by means of some representation results, we connect the study of finite closure systems, matroids and finite lattice theory in the scope of the more general notion of attribute dependency based on information tables. For a fixed finite set Ω and for a corresponding information table $\mathcal{J}$ having attribute set Ω, the fundamental tool we use to proceed in our investigation is the equivalence relation $\approx_{\mathcal{J}}$ on the power set $\mathcal{P}(\Omega)$ which identifies two any sets of attributes inducing the same indiscernibility relation on the object set of $\mathcal{J}$. We interpret the attribute dependency as a preorder $\geq_{\mathcal{J}}$ on $\mathcal{P}(\Omega)$ whose induced equivalence relation coincides with $\approx_{\mathcal{J}}$. Then we investigate in detail the links between the preorder $\geq_{\mathcal{J}}$, a closure system and an abstract simplicial complex on Ω naturally induced by $\approx_{\mathcal{J}}$ and specific families of ordered pairs of sets of attributes on Ω.

1. Introduction

In two classical papers [48,49], Zadeh introduced a new perspective in order to investigate fuzzy set theory, that he called *granular computing* (briefly GrC). In a short time, GrC became a field of research on its own [34]. From a very general viewpoint, GrC can be considered as a collection of all theories, methodologies, techniques and tools simulating the human skill to perceive the real world and to synthetize knowledge abstracting only the needed informations. Despite of its simplicity, GrC encloses a series of remarkable results and techniques that due to their generality can be used in several fields of discrete mathematics and computer science.

However, the basic idea underlying GrC is that the way in which we think of a given universe of objects leads us to a classification and grouping of such objects at various levels of detail. Hence, GrC methodologies try to provide formally precise criteria for moving from a more general level of detail to a more analytical one, and vice versa. On the other hand, a subsequent analysis of the various levels of detail of the above classifications of our universe can alter our initial thinking about objects and relationships between them. In this sense, we can say that GrC is closely related to the human way of thinking.

In this paper we investigate some mathematical structures which naturally arise when some ideas derived by GrC are applied on information tables (briefly GrC-IT). An *information table*, or equivalently an *information system* (see [18] for recent

* Corresponding author.
E-mail addresses: giampiero.chiaselotti@unical.it (G. Chiaselotti), gentile@mat.unical.it (T. Gentile), f.infusino@mat.unical.it (F. Infusino).

https://doi.org/10.1016/j.ins.2018.02.046
0020-0255/

studies related to the use of this double terminology in literature) is a quadruple $\mathcal{J} = \langle U, \Omega, F, \Lambda\rangle$, where U and Ω are two finite sets whose elements are respectively called *objects* and *attributes*, Λ is a non-empty set of elements called *values* and F is a map from $U \times \Omega$ into Λ called *information map*. As a matter of fact, information tables arise from the need of studying knowledge representation and data extraction. For example, in database theory there is a very frequent need of studying finite tables having a very large quantity of data, therefore many researches have been directed towards the purpose of reducing and simplifying the interpretation of these data.

A substantial part of GrC-IT is based on *rough set theory* (briefly RST), which originated at first as a methodology useful to extract and minimize deduction rules from data tables [33]. Next, the theoretical foundations of RST have been widely developed in connections with several areas: algebra [30,45], formal logic [7,31,44], fractal theory [37].

In a broader perspective, GrC paradigm [34] includes relevant parts of RST [47], probabilistic RST [40], rough mereology [36,38], database theory [20,21], data mining [26], graph and hypergraph theory [9,10,41] and parts of fuzzy set theory [35] within a more general perspective, in which complex systems are investigated grouping their elements into blocks called *granules*. Further types of interrelations between GrC and RST are described in [22].

However, at a first basic level, the classical granular notion of RST is the indiscernibility relation: by fixing an attribute subset in an information table, one arranges the objects in *information granules*, in such a way that all the objects in a given granule assume the same value on each single attribute of the fixed subset.

Starting with the notion of indiscernibility induced by an information table, in [6,11,14] some basic subfamilies of attribute subsets have been framed in a GrC context. In this perspective, some of these structures have been interpreted as *micro granular structures* and others as *macro granular structures* (for details see [6]). For decision tables, an analogous investigation has been initiated in [16]. In this paper we continue the investigation of the fundamental mathematical properties of the aforementioned micro and macro structures and the way they interact with each other.

In this work we assume that Ω is a fixed finite set, $\mathcal{P}(\Omega)$ is the power set of Ω and $\mathcal{J} = \langle U, F, \Lambda\rangle$ is an information table having Ω as attribute set (briefly, an information table on Ω). Then we will consider the equivalence relation $\approx_{\mathcal{J}}$ on $\mathcal{P}(\Omega)$ that identifies any two subsets of attributes $A, A' \in \mathcal{P}(\Omega)$ inducing on U the same indiscernibility relation [19,23,28,29,39]. In our context (see also [6,11]) we call $\approx_{\mathcal{J}}$ *indistinguishability relation* of $\mathcal{J}$.

This paper can be considered the natural continuation of the research project outlined in [6], which consists of investigating the mathematical foundations of the families of sets of attributes (possibly infinite) and related operators induced by the above indistinguishability relation. To this regard, we work on two levels: the interpretation of some parts of RST within an operator-based perspective and the development of new mathematical results concerning such operators and the associated families of subsets. The genesis of the ideas that we have largely used in this paper is due to Pagliani and Chakraborty [32] and to Polkowski [38]. In fact, many of the mathematical results obtained in this paper are a natural development of the conceptual geometric framework outlined in [32]. Moreover, the paradigm of mereology applied on RST (as described in [38]) provides a philosophical framework in which the mathematical interrelations between micro and macro structures of GrC-IT have been investigated and analyzed in [6] and in the present paper. Based on the previous considerations, to better illustrate the results and contributions of our work, we divide this introductory section in several sub-sections.

1.1. Basic facts on information tables

Let Ω be a fixed finite set, that we will consider as a given attribute set. Let $\mathcal{J} = \langle U, \Omega, F, \Lambda\rangle$ be an information table on Ω. For any subset $A \in \mathcal{P}(\Omega)$, we consider the usual *A-indiscernibility relation* $\equiv_A$ on U, defined by $u \equiv_A u'$ if $F(u, a) = F(u', a)$ for any $a \in A$ and for any u, $u' \in U$. Let $[u]_A$ be the equivalence class of u with respect to $\equiv_A$ and $\pi_{\mathcal{J}}(A) := \{[u]_A : u \in U\}$ the partition on U induced by $\equiv_A$, that here we call respectively *A-indiscernibility class* of u and *A-indiscernibility partition* of $\mathcal{J}$.

Let $PIND(\mathcal{J}) := \{\pi_{\mathcal{J}}(A) : A \in \mathcal{P}(\Omega)\}$ and $\mathbb{P}_{ind}(\mathcal{J}) := (PIND(\mathcal{J}), \preceq)$, where $\preceq$ is the usual refining partial order between partitions of U. In [11,23] it has been proved that $\mathbb{P}_{ind}(\mathcal{J})$ is a complete lattice, called the *indiscernibility partition lattice* of $\mathcal{J}$.

On the other hand, formally, the *indistinguishability relation* of $\mathcal{J}$ is the equivalence relation $\approx_{\mathcal{J}}$ on the power set $\mathcal{P}(\Omega)$ defined by

$$A \approx_{\mathcal{J}} A' :\iff \pi_{\mathcal{J}}(A) = \pi_{\mathcal{J}}(A'), \tag{1}$$

for any $A, A' \in \mathcal{P}(\Omega)$.

If $[A]_{\approx_{\mathcal{J}}}$ is the equivalence class of A in $\mathcal{P}(\Omega)$ with respect to $\approx_{\mathcal{J}}$, we call it the *$\mathcal{J}$-indistinguishability class* (or also *indistinguishability granule*) of A.

The indistinguishability relation has been formally introduced in [28,29,39] and further investigated in GrC-IT [6,11,14]. Moreover, this notion has been implicitly used even earlier to investigate graphs [10,12,13], hypergraphs [9], digraphs [15] and some types of discrete dynamical systems [1,5,8] (more in general, for temporal dynamical interpretations of RST see [17]).

The basic idea behind the introduction of the indistinguishability relation consists of focusing our attention on the attribute set Ω rather than on all the corresponding indiscernibility partitions induced on the object set U. To be more specific, whenever we investigate the partition of U induced by a given attribute subset A of Ω, we perform on the object set U a type of *local* granular computing, because it depends on the choice of A. On the other hand, when we partition the power set $\mathcal{P}(\Omega)$ by means of the indistinguishability classes, we obtain a type of *global* granular classification. Such a global