Information Sciences 330 (2016) 342–357

Contents lists available at ScienceDirect

Information Sciences

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

On the connection of hypergraph theory with formal concept analysis and rough set theory

Gianpiero Cattaneo [a], Giampiero Chiaselotti [b], Davide Ciucci [a,*], Tommaso Gentile [b]

[a] Department of Informatics, Systems and Communications, University of Milano-Bicocca, Milano 20126, Italy
[b] Department of Mathematics and Informatics, University of Calabria, Via Pietro Bucci, Cubo 30B, Arcavacata di Rende (CS) 87036, Italy

ARTICLE INFO

Article history:
Received 3 December 2014
Revised 4 August 2015
Accepted 30 September 2015
Available online 16 October 2015

Keywords:
Rough sets
Formal concept analysis
Hypergraphs

ABSTRACT

We present a unique framework for connecting different topics: hypergraphs from one side and Formal Concept Analysis and Rough Set Theory from the other. This is done through the formal equivalence among Boolean information tables, formal contexts and hypergraphs. Links with generic (i.e., not Boolean) information tables are established, through so-called nominal scaling. The particular case of k-uniform complete hypergraphs will then be studied. In this framework, we are able to solve typical problems of Rough Set Theory and Formal Concept Analysis using combinatorial techniques. More in detail, we will give a formula to compute the degree of dependency and the partial implication between two sets of attributes, compute the set of reducts and define the structure of the partitions generated by all the definable indiscernibility relations.

1. Introduction

Hypergraph theory, Formal Context Analysis and Rough Set Theory are three well-developed fields of study.

Hypergraph theory is a generalization of graph theory (see [4,5]) where edges, called *hyperedges*, can have an arbitrary number of vertices. Classically, the typical problems studied in hypergraph theory concern combinatorial questions (see [[5], [8]]) and optimization questions (see [20]).

Formal Context Analysis (briefly FCA) deals with structures called *Formal Contexts* [17], that describe objects in terms of the properties they possess. A *formal concept* is an objects–properties pair (O, P) such that the objects in O are all and the only to satisfy all the properties in P. Starting then from the definition of formal concept, the FCA theory developed as a very powerful theoretical methodology useful to approach problems in data mining, machine learning and related fields (see [21]).

Finally, an *Information System* (or information table) collects the values that a set of objects have on some attributes. Objects are then partitioned according to an equivalence relation in classes of objects with equal values for all attributes. The pair (U, R) made of the objects and the equivalence relation is called *approximation space*, and it can be analyzed by using the methods derived from Rough Set Theory (RST) (see [26,27]).

In particular, an Information System is called *Boolean* if we fix the set of attribute values equal to $\{0, 1\}$ and it is easily proved to be equivalent to a formal context. See for instance [35–37], where a unifying approach between FCA theory and Boolean

* Corresponding author. Tel.: +39 02 64487902; fax: +39 02 64487839.
E-mail addresses: cattang@live.it (G. Cattaneo), giampiero.chiaselotti@unical.it (G. Chiaselotti), ciucci@disco.unimb.it, ciucci@disco.unimib.it (D. Ciucci), gentile@mat.unical.it (T. Gentile).

http://dx.doi.org/10.1016/j.ins.2015.09.054

information system theory has been outlined. Let us notice that several authors studied the link between RST and FCA and mixed the two theories in several ways, see for instance [18,19,25,31,32].

On the other hand, in several other papers (see [14,15,29,30]) specific hypergraphs associated with some particular type of formal contexts and Boolean information tables were constructed. For instance, Stell applies rough set and formal concept ideas to hypergraphs: in [29] rough hypergraph theory is introduced and, similarly, in [30] FCA is generalized using a relation on hypergraphs instead of sets. In [14] a hypergraph is considered as a basic model for granular structures. Results in [15] directly put in relation a formal context with a hypergraph but differently as we do (see Remark 3.9).

In this paper, we survey some results concerning the relationship among formal contexts, Boolean information tables and hypergraphs. Moreover, through a scaling procedure we see that information tables (not necessarily Boolean) are in relation with a sub-class of hypergraphs. Several consequences are then derived from this unified framework.

First of all, in the next example we show how some questions concerning a typical problem derived from an information table can be naturally interpreted as a combinatorial problem in hypergraph theory.

Example 1.1. Let us consider an information table with 100 distinct attributes $Att = \{a_1, \ldots, a_{100}\}$ and suppose that we can uniquely characterize any object by exactly 7 attributes (not necessarily the same for all objects). For instance, the sequence of attributes

$$a_1 a_3 a_4 a_{15} a_{23} a_{71} a_{98}$$

uniquely determine a specific object of our universe and, on the other hand, each object of our universe is uniquely determined by a "code" of length 7 of distinct attributes selected from $a_1, \ldots, a_{100}$. In this case, we can represent such a situation as a new Boolean information system, where the attributes are $Att := \{a_1, \ldots, a_{100}\}$, the objects are all the 7-subsets $u, v, \ldots$ of Att and the value of an object u on an attribute a is equal to 1 if $a \in u$ and equal to 0 otherwise. We choose now a particular subset A of Att, such that A has at most seven elements. For example $A = \{a_1, a_2, a_{50}, a_{51}, a_{97}\}$. Let us suppose that a user wants to know the following information:

1. How many distinct equivalence classes of items with respect to attributes A are there?
2. Which are the items having as part of their code the sub-string $a_2 a_{50} a_{97}$, but none of the symbols a_1, a_{51}?

The previous can be considered two very practical requests for a user, and they can be understood more effectively after being framed in Information System theory, by using combinatorial techniques (see Remark 4.5 and Proposition 4.6). Let us note that, in terms of Information Systems, the request (2) is equivalent to determine all the objects in our universe that are in a specific A-indiscernibility class, namely those with attributes $a_2 a_{50} a_{97}$ equal to 1 and attributes $a_1 a_{51}$ equal to 0 in the original information table.

The previous example shows that it can be useful to define and to study some combinatorial properties of Information Systems where we have a fixed set of attributes $a_1, a_2, \ldots$ and where each object of the universe is uniquely determined by a choice of distinct attributes. Hence, each object u can be identified with some subset of $a_1, a_2, \ldots,$ and this situation can be represented by a hypergraph, where attributes are the vertices and objects are the hyperedges. Mathematical literature (see [[4,5], [8], [20]]) provides several examples of hypergraph families. A classical example of deeply studied hypergraph family are (n,k) uniform complete hypergraphs $\{\binom{\hat{n}}{k} : n \geq k \geq 0\}$, where $\binom{\hat{n}}{k}$ is the set of all subsets with k elements of the n-set $\hat{n} := \{1, \ldots, n\}$. This family of hypergraphs is exactly the one needed to answer questions (1) and (2) above.

Moreover, as another example of the possibility given by the connection among information tables, formal contexts and hypergraphs, we will study the notion of attribute dependency with the help of k-uniform complete hypergraphs. In particular we will be able to compute the RST-dependency degree of a set of attributes B with respect to a set of attributes A and the FCA-precision of a partial implication [23].

The paper is organized as follows. In Section 2, we give the basic definitions of the three involved theories: Rough Set Theory (in particular, information tables, approximations and attribute dependence), formal context analysis, hypergraphs. The equivalence of the three involved structures to represent data is then given in Section 3.1. In Section 4 we study in detail our "hypergraph model" $\binom{\hat{n}}{k}$ and we apply on this specific model all the general theoretical tools introduced in the previous sections. In Section 5 we define the notion of granular partition lattice for a hypergraph and we determine this lattice for the hypergraph $\binom{\hat{n}}{k}$. Moreover, we establish an isomorphism between the granular partition lattice and another lattice derived from pattern structures, as introduced in [16]. Finally, in Section 6 we draw some conclusions and outline future works.

2. Basic notions

In this section, the preliminary notions of Rough Set Theory, Formal Concept Analysis and Hypergraphs are introduced.

2.1. Rough Set Theory

An *Information System* is a structure $\mathcal{I} = \langle U, Att, V, F \rangle$, where U (called *universe set*) is a non-empty set of *objects*, Att (called *attribute set*) is a non-empty set of *attributes*, V (called *value set*) such that $V = \bigcup_{a \in Att} V_a$ where V_a is the set of values that attribute a can assume and each V_a is non-empty and F: $U \times Att \to V$ (called *information map*) is a mapping from the direct product $U \times Att$ into the value set V.