Computer Science and Information Systems 11(2):779–796 DOI:10.2298/CSIS130125042G

Approximation to the Theory of Affinities to Manage the Problems of the Grouping Process

Anna M. Gil-Lafuente[1] and Anna Klimova[1]

[1] Department of Business Administration, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
{amgil, aklimova}@ub.edu

Abstract. New economic and enterprise needs have increased the interest and utility of the methods of the grouping process based on the theory of uncertainty. A fuzzy grouping (clustering) process is a key phase of knowledge acquisition and reduction complexity regarding different groups of objects. Here, we considered some elements of the theory of affinities and uncertain pretopology that form a significant support tool for a fuzzy clustering process. A Galois lattice is introduced in order to provide a clearer vision of the results. We made an homogeneous grouping process of the economic regions of Russian Federation and Ukraine. The obtained results gave us a large panorama of a regional economic situation of two countries as well as the key guidelines for the decision-making. The mathematical method is very sensible to any changes the regional economy can have. We gave an alternative method of the grouping process under uncertainty.

Keywords: fuzzy logic, clustering process, affinities, pretopology, Galois Lattices.

1. Introduction

The intelligibility of the universe for each person depends on his aptitude to group or classify different objects. Identification of object types is one of the first phases of knowledge acquisition [1]. “Organizing data into sensible groupings is one of the most fundamental modes of understanding and learning” [2]. “A fundamental operation in data mining is the partitioning of a set of objects represented by data into homogeneous groups or clusters” [3].

The analysis of the process of groups making (clustering analysis) is fundamental nowadays for understanding the complex processes of various substantive areas such as medicine [4-9], chemical industry [10-12], engineering [13-16], image processing [17-19], business [20-35] and others.

At the beginning Boolean logic has appeared as one of the most powerful mathematical tools able to make different groups giving the adequate solutions [36]. It has been commonly used in the classical clustering analysis methods that are the most important techniques in the information procedure [1]. Clustering has been extensively studied in machine learning, databases, and statistic from various perspectives. There are two classical hard clustering methods. Statistical clustering methods [37-39] partition objects according to some proximity (similarity or dissimilarity) measures, whereas conceptual clustering methods [40,41] cluster objects according to the concepts the objects have. For automatic clustering, a method of determining proximity between feature vectors as well as a method for determining representatives of clusters is required. Hard clustering method is more adequate for clustering condition with clear boundary and preciseness in databases [1].

Although a basis of classical clustering techniques remains the same, some changes have been recently introduced into their structure. With increasing complexity of the technological and economic processes, an important aspect of cluster analysis comes to the fore which is the rigidity of the partition that is necessary to find. Many authors have proposed a fuzzy setting as the adequate approach to deal with this problem. Fuzzy set theory, introduced by Zadeh [42], gives an idea of uncertainty of belonging, which is described by a membership function. Fuzzy set theory represents a new model of grouping process that allows classifying the elements under uncertainty. During the last decades, extensive research has been carried out with respect to the investigation of fuzzy clustering techniques for classification. See Refs. 43-54.

In the early 90s a new approach to fuzzy clustering process was proposed by some authors [33].The concept of affinity between the elements has been considered as an “angular stone” of segmentation process under uncertainty. The theory of the affinities represents a generalization of the relations of similarity extending the field of their performance to the field of the rectangular matrices.

The notion of similarity and affinity represent different ways of expressing the concept of neighborhood. Similarity indicates a specific resemblance, either partial or total, between two physical or imaginary objects [51-52]. Affinity refers to a collective behavior of objects with respect to certain specific criteria, even when these criteria are not particularly well specified. But both of these notions indicate whether two or more objects are related under adequate conditions with respect to explicit criteria for affinity [45].

In this paper we have proposed an extension of the algorithm based on the theory of affinities for the formation of homogeneous groups of economic regions of Russian Federation and Ukraine. With this in mind, we have pretended to offer an alternative solution to the problems faced by every person in charge of the distribution (investment) of company financial resources. It was essential, for us, to make a good economic and financial assessment of the regions of Russia Federation and Ukraine, providing a wide and faithful basis for making-decision process under uncertainty.

In this study we have also considered some elements of uncertain “pretopology” that have served us as a mathematical support to implement the theory of affinities to the practical case. See Refs. 23,26,55-58. Some concepts of combinatorial analysis [59] have been introduced too. As a culminating element of our paper we have used a Galois lattice. We considered it an effective tool to structure the obtained results.

2. Recent Backgrounds

In a previous part we have already told about general tendencies of the development of the mathematical methods and algorithms related to the grouping/ clustering process. Now let us continue briefly with the recent approaches that are closely related to the present paper and represent special interest for the current topic.

Gil Aluja in his paper titled “Clans, affinities and Moore’s fuzzy pretopology” [60] proposed some new elements that renovate the structures of the economic thought based on the geometric idea. The elements are based on the concept of pretopology. The author considered that it was impossible to conceive, today, formal structures able to represent the Darwininan conception of the economic behavior without appealing to this part of combinatory mathematics. The author proposed the suitable schemes for the treatment of the economic and management phenomenon using the theory of clans, affinities and the Moore’s closures.

In one of his next papers Gil Aluja emphasized that very often in situations of uncertainty in portfolio management it is difficult to apply the numerical methods based on the linearity principle. In this case, nonnumeric techniques were used to assess the situations with a nonlinear attitude. The authors proposed to use a concept of grouping providing, by this manner, good solutions to the problems of the homogeneous groupings. The Pichat and affinities algorithms were proposed to reduce the number of elements of the power sets of the titles listed in the Stock Exchange and to group them correctly [61].

Among other recent works represented by the same authors dedicated to the main topic of this paper we can mention the following “Decision-making techniques with similarity measures and OWA operators” [62]. In the last paper several aggregation techniques such as the Hamming distance, the adequacy coefficient and the index of maximum and minimum level that were very useful for decision-making were considered. These were the useful mathematical tools able to determine similarities between different elements of the system helping in decision making process especially in business and economic areas (in production management in this concrete case).

The article titled “Approaches to managing sustainability among enterprises” [63] analyzed the important changes a business environment where the companies perform experienced during the last decades. It was characterized as uncertain and unstable. Basing on the idea of the importance and complexity the sustainability management represents for the companies the authors proposed the use of flexible tools based on fuzzy logic such as the Theory of Affinities, the Clans Theory, Hungarian Algorithm, etc., in order to assist the enterprises in making decisions and help them to improve management with stakeholders contributing to the treatment of the problems in the future.

Another field of the social sciences where the algorithms and technics based on the mathematics of uncertainty were used to find out the coherent solutions for decision making was a human resource management. The purpose of the article titled “A personal selection model using Galois group theory” was to propose a personal selection model based on the comparison between the qualifications of prospective candidates. The model based on the Galois group theory was used to group different candidates depending on the common characteristics that meet a certain level. See Ref. 64.

In particular, another author proposed to use the Pichat algorithm, among other methods, in the sports area. He demonstrated that this tool was a great help in decision making related to the phenomenon of grouping/ clustering. This algorithm showed whether the members of a human group/a sport team were sufficiently and properly interrelated or not and, as a consequence, whether its performance was optimal or not. See Ref. 65.

Because of the limitations of the article text extension we'll finish our explications here. We hope that this brief analysis showed the importance and usefulness of the grouping methods based on the mathematics of uncertainty in different social sciences nowadays especially in economy and management.

3. The Theory of the Affinities

The theory of the affinities represents a generalization of the relations of similarity extending the field of their performance to the field of the rectangular matrices. This theory is applied in a multitude company's management problems.

The attempt to generalize a notion of similarity was initiated by Jaime Gil Aluja and Arnold Kaufmann in the eighties. The theory of the affinities turned out to be a successful result of the work presented by these scientists at the IX European Congress of the Operative Investigation.

"We determine the affinities as those homogeneous orderly structured groups limited to the established levels which join all elements of two sets of different nature. These two sets are related between themselves by the own essence of the phenomenon which represent them" [33].

There are three main aspects which form the concept of affinity. The first one refers to the fact that the homogeneity of each group is related to the established level. According to the requirement of each characteristic (the element of one of the sets) the high level of a determiner is assigned and that is a threshold from which the homogeneity starts to exist. The second aspect shows a necessity of the fact that the elements of each set are to be connected between themselves by the certain rules of the nature or human will. The third aspect requires creating a structure which has a special order that is capable to be involved into the process of the decision making.

Over the last decades many approaches to the theory of affinities have been proposed and several useful techniques based on this theory have been provided. See Refs. 23-32,45,51,52,58,66-68.

4. Axioms for the Uncertain Pretopology

There are two elements such as a finite referential E and a "power set" $P(E)$ that have a different nature. Let us define uncertain pretopology. See Refs. 26,33,58.

Definition 1. A mapping Γ of $P(E)$ in $P(E)$ is uncertain pretopology of E if, and only if, the following axioms are given:

1) $\Gamma\varnothing = \varnothing$

2) $\forall \underset{\sim}{A}_k \in P(E) : \underset{\sim}{A}_k \subset \Gamma \underset{\sim}{A}_k$

(where $\Gamma E = E$ and $\underset{\sim}{A}_k$ is a fuzzy subset)

Using the same terms as in determinism, a mapping Γ for all elements of $P(E)$ will be called as an "adherent mapping" or "adherence". It is also possible to associate Γ with an "inner mapping" or "inner" δ, defined by:

$$\forall \underset{\sim}{A}_k \in P(E): \ \delta \underset{\sim}{A}_k = \overline{\Gamma \overline{\underset{\sim}{A}_k}} \tag{1}$$

(where $\overline{\underset{\sim}{A}_k}$ is a complement of $\underset{\sim}{A}_k$).

Definition 2. $\underset{\sim}{A}_k$ is closed set for an uncertain pretopology if $\underset{\sim}{A}_k = \Gamma \underset{\sim}{A}_k$.

Definition 3. $\underset{\sim}{A}_k$ is open set for an uncertain pretopology if $\underset{\sim}{A}_k = \delta \underset{\sim}{A}_k$.

We considered the widest notion of uncertain pretopology. If new axioms are added then new uncertain pretopological spaces appear. They have a special interest for the treatment of the problems of different economic systems.

Definition 4. A mapping Γ of $P(E)$ in $P(E)$ is uncertain isotone pretopology of E if, and only if, the following axioms are carried out:

1) $\Gamma\varnothing = \varnothing$
2) $\forall \underset{\sim}{A}_k \in P(E) : \underset{\sim}{A}_k \subset \Gamma \underset{\sim}{A}_k$, extensivity
3) $\forall \underset{\sim}{A}_k, \underset{\sim}{A}_l \in P(E) : (\underset{\sim}{A}_k \subset \underset{\sim}{A}_l) \Rightarrow (\Gamma \underset{\sim}{A}_k \subset \Gamma \underset{\sim}{A}_l)$, isotony

Let us pay a special attention to one of the uncertain isotone pretopologies that satisfies the property, which is idempotency, and can be represented as the fourth axiom.

4) $\Gamma(\Gamma \underset{\sim}{A}_k) = \Gamma \underset{\sim}{A}_k$, idempotency

With these four axioms, we get Moore's closure which, moreover, has the property $\Gamma\varnothing = \varnothing$, not required in the axiomatic of this closure.

Moore's closure is of a special importance in the development of the algorithms capable of dealing with segmentation problems. So it is necessary to establish ways for obtaining Moore's closures on the basis of the concepts that could easily be found in the information supported by different institutions.

5. Obtaining a Moore's Closure

On the basis of previous practice, we have found two concepts that seem to adapt well to the exposed necessities. These are the notion of a Moore's family and a graph or fuzzy relation. See Refs. 33,51,52.

- Axiomatic of a Moore's family.

Definition 5. Let one family be determined through $F(E) \subset P(E)$. If there are two axioms:

1) $E \in F(E)$
2) $(\underset{\sim}{A}_k, \underset{\sim}{A}_l \in F(E)) \Rightarrow (\underset{\sim}{A}_k \cap \underset{\sim}{A}_l \in F(E)),$

then this is a Moore's family.

Moore's closure is obtained by the following way:

1) Once the family $F(E)$ is available it is necessary for each element $\underset{\sim}{A}_k \in P(E)$ to find a subset of elements of $F(E)$ that contain $\underset{\sim}{A}_k$ and will be referred to as $F_{\underset{\sim}{A}_k}(E)$.
2) For each $F_{\underset{\sim}{A}_k}(E)$ its intersection $\bigcap_{F \in F_{\underset{\sim}{A}_k}(E)} F$ is obtained.
3) A mapping $M\underset{\sim}{A}_k$ is determined by the mode that $\bigcap_{F \in F_{\underset{\sim}{A}_k}(E)} F$ will be corresponded to each element $\underset{\sim}{A}_k$, that is to say: $M\underset{\sim}{A}_k = \bigcap_{F \in F_{\underset{\sim}{A}_k}(E)} F$.

So, Moore's closure was found using one family. There is only one closure for each family.

- The concept of relation is at the basis of a majority of algorithms of non-numerical mathematics of uncertainty, considered by Zadeh [54] and Kaufmann [51]. To formalize this relation there are two graphs that can be used represented either in matrix form or arrowlines.

  Let us consider a fuzzy relation $[\underset{\sim}{R}]$ between the elements of a referential $E = \{y_1, y_2, ..., y_n\}$, such as $[\underset{\sim}{R}] \subset E \times E$.

  1) With this information a Boolean graph is determined, taking a threshold α as follows:

$$[R_\alpha] = \{(y_i, y_j) \subset E \times E \,/\, \mu_{\underset{\sim}{R}}(y_i, y_j) \geq \alpha\}, \alpha \in [0,1]. \qquad \textbf{(2)}$$

  2) Predecessors and successors of y_i limited to the level α are determined as:

$$\left.\begin{array}{l} y_j \text{ is a sucessor of } y_i \\ y_i \text{ is a predecessor of } y_j \end{array}\right\} \text{ if } (y_i, y_j) \in [R_\alpha]. \tag{3}$$

These concepts allow us to find the connection to the right $R_\alpha^+ \underset{\sim}{A}_k$ and to the left $R_\alpha^- \underset{\sim}{A}_k$.

Definition 6. A connection to the right $R_\alpha^+ \underset{\sim}{A}_k$ or, a mapping R_α^+ of $P(E)$ in $P(E)$ such that for all $\underset{\sim}{A}_k \in P(E)$, R_α^+ is a subset of elements of E that are successors, limited to the level α, of all element that belong to $\underset{\sim}{A}_k$, expressed as:

$$R_\alpha^+ \underset{\sim}{A}_k = \{y_j \in E / (y_i, y_j) \in [R_\alpha], \forall\, y_i \in \underset{\sim}{A}_\alpha\}, \tag{4}$$
$$\text{with } R_\alpha^+ \varnothing = E.$$

Definition 7. A left connection $R_\alpha^- \underset{\sim}{A}_k$ or, a mapping R_α^- of $P(E)$ in $P(E)$ such that for all $\underset{\sim}{A}_k \in P(E)$, R_α^- is a subset of the elements of E that are predecessors, limited to the level α, of all elements that belong to $\underset{\sim}{A}_k$. It is expressed as:

$$R_\alpha^- \underset{\sim}{A}_k = \{y_i \in E / (y_i, y_j) \in [R_\alpha], \forall\, y_j \in \underset{\sim}{A}_\alpha\}, \tag{5}$$
$$\text{with } R_\alpha^- \varnothing = E.$$

3) The right connection $R_\alpha^+ \underset{\sim}{A}_k$ and the left one $R_\alpha^- \underset{\sim}{A}_k$ can be found directly by simple reading of the relation $[R_\alpha]$, in the following way:

$$\forall\, y_i \subset \underset{\sim}{A}_k \in P(E):\ R_\alpha^+ \underset{\sim}{A}_k = \bigcap_{y_i \subset \underset{\sim}{A}_k} R_\alpha^+ \{y_i\}, \tag{6}$$

$$R_\alpha^- \underset{\sim}{A}_k = \bigcap_{y_i \subset \underset{\sim}{A}_k} R_\alpha^- \{y_i\}. \tag{7}$$

4) Two Moore's closures corresponding to the fuzzy relation $[R_\alpha]$ or to the graph, limited to level α, are obtained by the maxmin convolution R_α^- with R_α^+ and R_α^+ with R_α^- . It is written:

$$M_\alpha^{(1)} = R_\alpha^- \bullet R_\alpha^+ \text{ and } M_\alpha^{(2)} = R_\alpha^+ \bullet R_\alpha^-, \tag{8,9}$$

where $M_\alpha^{(1)}$ and $M_\alpha^{(2)}$ are the two Moore's closures.

Two concepts, such as a family and a graph, can be used to obtain Moore's closures which are considered to be important for a wide range of economic approaches.

6. Moore's Closed Sets as Maximum Groups

As for any pretopology, a set of Moore's closed sets is formed by the elements that comply with the following condition:

$$\forall \underset{\sim}{A}_k \in P(E): \underset{\sim}{A}_k = M\underset{\sim}{A}_k. \quad \mathbf{(10)}$$

Let us determine a subset of closed sets of $P(E)$ corresponding to Moore's closure $M_\alpha^{(2)}$ as $C(E, M_\alpha^{(2)})$. As $R_\alpha^- \underset{\sim}{A}_k$ is a closed of $P(E)$ for $M_\alpha^{(2)}$ which can be written as follows:

$$C(E, M_\alpha^{(2)}) = \bigcup_{\underset{\sim}{A}_k \in P(E)} R_\alpha^- \underset{\sim}{A}_k. \quad \mathbf{(11)}$$

and, $R_\alpha^+ \underset{\sim}{A}_k$ is a closed for $M_\alpha^{(1)}$ and we determine the subset of closed sets of $P(E)$ corresponding to Moore's closure $M_\alpha^{(1)}$ as $C(E, M_\alpha^{(1)})$, we get:

$$C(E, M_\alpha^{(1)}) = \bigcup_{\underset{\sim}{A}_k \in P(E)} R_\alpha^+ \underset{\sim}{A}_k. \quad \mathbf{(12)}$$

Both families $C(E, M_\alpha^{(1)})$ and $C(E, M_\alpha^{(2)})$ are isomorphic and dual for each other. They are also antitone. Therefore, we get:

$$\underset{\sim}{A}_k \in C(E, M_\alpha^{(1)}) \Rightarrow (\underset{\sim}{A}_l = R_\alpha^- \underset{\sim}{A}_k \in C(E, M_\alpha^{(2)}) \text{ and } R_\alpha^+ \underset{\sim}{A}_l = \underset{\sim}{A}_k). \quad \mathbf{(13)}$$

$$\underset{\sim}{A}_l \in C(E, M_\alpha^{(2)}) \Rightarrow (\underset{\sim}{A}_k = R_\alpha^+ \underset{\sim}{A}_l \in C(E, M_\alpha^{(1)}) \text{ and } R_\alpha^- \underset{\sim}{A}_k = \underset{\sim}{A}_l). \quad \mathbf{(14)}$$

These families of closed sets can be associated to each other and each of the families forms a finite lattice [51]. As a result of the maxmin convolution, both families of closed sets provide the groupings with the greatest possible number of elements of the referential E. In the formation of groups, the groupings of the elements of one family of the closed sets are accompanied by the other groupings of the other family of closed sets related to them.

7. Galois Lattice and its Importance in the Process of Segmentation

There is a fuzzy rectangular relation $[\underset{\sim}{R}] \subset E_1 \times E_2$. In this case the necessary properties obtained for only one referential are kept. See Refs. 29-33. Only one lattice is obtained after two lattices were joined, where each vertex represents the relation of the groupings of elements of set E_1 with the groupings of elements of set E_2 taking into consideration a previously established level α. If this happens then there is affinity between the groups of the elements limited to the correspondent level. This lattice determines a structure for these relations. See Refs. 25,27,28,32,33,56,57.

If, when the upper end of the lattice is different to (E_2, $\varnothing$) and the lower end is different to ($\varnothing$, E_1), we add these vertices, then we have a Galois lattice that shows up two sets of Moore's closed sets that have the previous relations (E_2,$\varnothing$) and ($\varnothing$, E_1) as the upper and lower ends.

Let us consider the relation of the following order:

$$\forall X, X' \in E^{(1)} \quad \forall Y, Y' \in E^{(2)}.$$

$$(X, Y \le (X', Y')) \leftrightarrow (X \subset X', Y \supset Y'). \quad \mathbf{(15)}$$

and, ∇ will be taken as an upper extreme of the lattice with the order of the Eq. (15).

By the same manner, the relation of the order:

$$\forall X, X' \in E^{(1)} \quad \forall Y, Y' \in E^{(2)}.$$

$$(X, Y \geq (X', Y')) \leftrightarrow (X \supset X', Y \subset Y'). \qquad \textbf{(16)}$$

will be associated with Δ which is a lower extreme of a Galois lattice (we deal with the complementary order).

If the following condition carries out:

$$(U, V) = (X, Y) \nabla (X', Y'). \qquad \textbf{(17)}$$

$$\text{then } (U \supset X \cup X' \text{ and } Y \subset Y \cap Y'). \qquad \textbf{(18),(19)}$$

The same situation happens if :

$$(Z, T) = (X, Y) \Delta (X', Y'). \qquad \textbf{(20)}$$

$$\text{then } (Z \subset X \cap X' \text{ and } T \supset Y \cup Y'). \qquad \textbf{(21),(22)}$$

Taking into account the Eqs. (3) to (8) we observe that the maximum subrelations have the configuration of a Galois lattice and the maximum subrelations of the similarity also have this configuration (they are not ordered between themselves in the partial order of the lattice, but they are ordered in the relation with the extremes ∇ and Δ).

8. Illustrative Example

We have participated in the process of formation of economic homogeneous groups of the regions of Russian Federation and Ukraine.

We will start our example with Russian Federation. We selected 7 regions and their 6 more significant social and economic attributes. There are two sets of the regions and their indicators:

$$E_1 = \{a, b, c, d, e, f, g\} \text{ and } E_2 = \{A, B, C, D, E, F\}.$$

So we have the following relation:

Table 1. Relations between two subsets.

	A	B	C	D	E	F
a	75.328	14.620	10.895	8	58.301	8.436

b	244.352	38.362	58.407	34	100.218	10.553
c	254.965	81.399	61.400	64	319.501	9.630
d	125.971	58.760	31.771	32	121.827	7.252
e	746.141	37.033	86.969	117	401.294	14.243
f	148.447	42.607	35.523	38	213.552	10.317
g	11.344	13.835	22.202	29	41.872	9.389

The data of this matrix is normalized. The average values for each column of the indicators are calculated. These values are considered as the presumption levels of homogeneity α from which the homogenous groups are determined. The level for A is 0,308, for B - 0,503, for C - 0,505, for D - 0,393, for E - 0,447 and for F - 0,700. So, the fuzzy relation is turned to a Boolean matrix, see below:

	A	*B*	*C*	*D*	*E*	*F*
a	0	0	0	0	0	0
b	1	0	1	0	0	1
c	1	1	1	1	1	0
d	0	1	0	0	0	0
e	1	0	1	1	1	1
f	0	1	0	0	1	1
g	0	0	0	0	0	0

Fig. 1. Boolean matrix

The right connection B^+ and the left connection B^- are established.

$$B^+\varnothing = E_2,\ B^+\{a\} = \varnothing,\ B^+\{b\} = \{A, C, F\},\ B^+\{c\} = \{A, B, C, D, E\}, ..,$$
$$B^+\{a, b, c, d, f\} = \varnothing, ..,\ B^+\{a, b, c, d, e, f, g\} = \varnothing.$$
$$B^-\varnothing = E_1,\ B^-\{A\} = \{b, c, e\},\ B^-\{B\} = \{c, d, f\},\ B^-\{C\} = \{b, c, e\}, ..,$$
$$B^-\{A, B, C, D, E\} = \{c\}, .., B^-\{A, B, C, D, E, F\} = \varnothing.$$

Then, Moore's closures such as $M^{(1)} = B^- \circ B^+$ and $M^{(2)} = B^+ \circ B^-$ are obtained.

In the next step the families of closed sets relative to the Moore's closures $M^{(1)}$ and $M^{(2)}$ are formed:

$$f(E, M^{(1)}) = \{\varnothing, \{B\}, \{E\}, \{F\}, \{A, C\}, \{B, E\}, \{E, F\}, \{A, C, F\}, \{B, E, F\}, \{A, C, D, E\},$$
$$\{A, B, C, D, E\}, \{A, C, D, E, F\}, E_2\}.$$

$$f(E, M^{(2)}) = \{\varnothing, \{c\}, \{e\}, \{f\}, \{b, e\}, \{c, e\}, \{c, f\}, \{e, f\}, \{b, c, e\}, \{b, e, f\},$$
$$\{c, d, f\}, \{c, e, f\}, E_1\}.$$

The families of closed sets which represent the isomorphic lattices are associated to each other. See Fig.2. A Galois lattice is represented at the Fig.3.

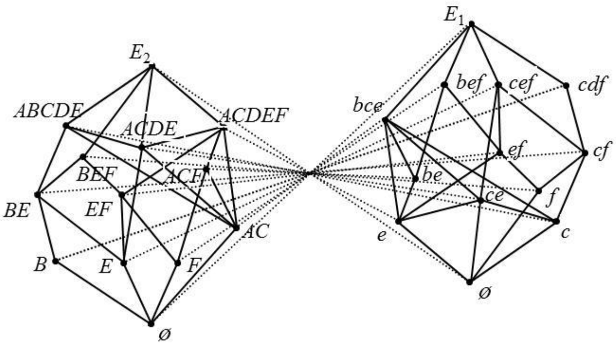

Fig. 2. Isomorphic lattices

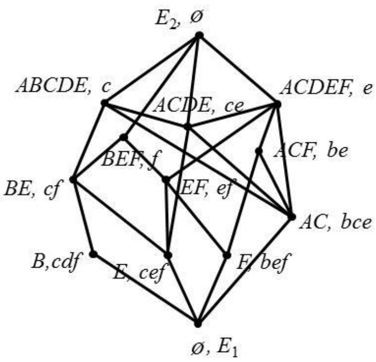

Fig. 3. Galois lattice

The following similar groupings of elements of two sets which are the set of regions and the set of attributes have been obtained.

$$\varnothing \leftrightarrow E_2,\ \{c\} \leftrightarrow \{A,B,C,D,E\},\ \{e\} \leftrightarrow \{A,C,D,E,F\},\ \{f\} \leftrightarrow \{B,E,F\},$$
$$\{b,e\} \leftrightarrow \{A,C,F\},\ \{c,e\} \leftrightarrow \{A,C,D,E\},...,\{b,c,e\} \leftrightarrow \{A,C\},\ \{b,e,f\} \leftrightarrow \{F\},$$
$$\{c,d,f\} \leftrightarrow \{B\},\ \ \{c,e,f\} \leftrightarrow \{E\},\ E_1 \leftrightarrow \varnothing.$$

Now we will do the same process of calculus for the economic regions of Ukraine. In this case we selected 6 regions and their 5 more significant economic attributes. There are two sets of the regions and their indicators:

$$E_1 = \{a,b,c,d,e,f\} \text{ and } E_2 = \{A,B,C,D,E\}.$$

So we have the following relation:

Table 2. Relations between two subsets.

	A	*B*	*D*	*E*	*F*
a	9.147	30.057	2.148	168	1.72
b	8.040	29.613	1.967	333	1.62
c	3.461	13.539	1.828	336	1.43
d	6.310	22.650	2.908	75	1.52
e	2.068	7.969	779	109	1.18
g	1.149	2.850	652	28	945

The data of this matrix is normalized, see below:

	A	*B*	*C*	*D*	*E*
a	1	1	0,7	0,4	1
b	0,8	0,9	0,6	0,9	0,9
c	0,3	0,4	0,6	1	0,8
d	0,6	0,7	1	0,2	0,8
e	0,2	0,2	0,2	0,3	0,6
f	0,1	0,0	0,2	0,0	0,5

Fig. 4. Normalized matrix

The average values for each column of the indicators are calculated. These values are considered as the presumption levels of homogeneity α from which the homogenous groups are determined. The level for *A* is 0,550, for *B* - 0,592, for *C* - 0,590, for *D* – 0,520 and for *E* - 0,814. So, the fuzzy relation is turned to a Boolean matrix, see below:

	A	*B*	*C*	*D*	*E*
a	1	1	1	0	1
b	1	1	1	1	1
c	0	0	1	1	1
d	1	1	1	0	1
e	0	0	0	0	0
f	0	0	0	0	0

Fig. 5. Boolean matrix

The right connection B^+ and the left connection B^- are established.

$$B^+\varnothing = E_2,\ B^+\{a\} = \{A,B,C,E\},\ B^+\{b\} = E_2, \ldots, B^+\{a,b\} = \{A,B,C,E\},$$

$$B^+\{a,c\} = \{C,E\}, \ldots, B^+\{c,d,e\} = \varnothing,\ B^+\{c,d,f\} = \varnothing,\ B^+\{c,e,f\} = \varnothing, ..,$$

$$B^+\{a,b,c,d,e\} = \varnothing,\ B^+\{a,b,c,d,f\} = \varnothing, \ldots,\ B^+\{a,b,c,d,e,f\} = \varnothing.$$

$$B^-\varnothing = E_1,\ B^-\{A\} = \{a,b,d\},\ B^-\{B\} = \{a,b,d\}, \ldots, B^-\{C,D\} = \{b,c\},$$

$$B^-\{C,E\} = \{a,b,c,d\}, \ldots, B^-\{B,C,D\} = \{b\},\ B^-\{B,C,E\} = \{a,b,d\}, \ldots$$

$$B^-\{B,C,D,E\} = \{b\},\ B^-\{A,B,C,D,E\} = \{b\}.$$

Then, Moore's closures such as $M^{(1)} = B^{-} \circ B^{+}$ and $M^{(2)} = B^{+} \circ B^{-}$ are obtained.

In the next step the families of closed sets relative to the Moore's closures $M^{(1)}$ and $M^{(2)}$ are formed:

$$f(E, M^{(1)}) = \{\varnothing, \{C,E\}, \{C,D,E\}, \{A,B,C,E\}, E_2\}.$$

$$f(E, M^{(2)}) = \{\{b\}, \{b,c\}, \{a,b,d\}, \{a,b,c,d\}, E_1\}.$$

The families of closed sets which represent the isomorphic lattices are associated to each other. See Fig. 6. A Galois lattice is represented at the Fig. 7.

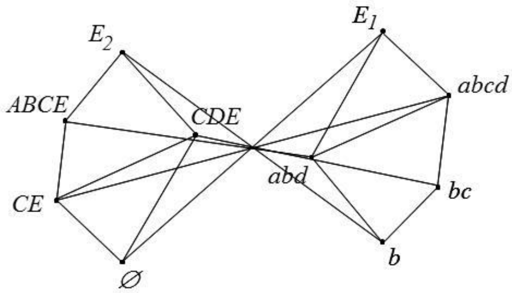

Fig. 6. Isomorphic lattices

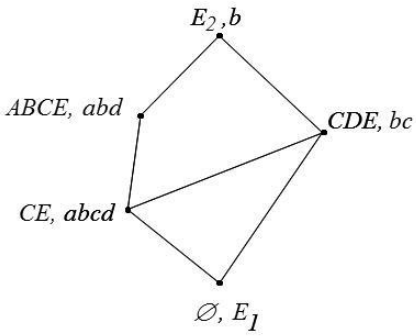

Fig. 7. Galois lattice

The following similar groupings of elements of two sets which are the set of regions and the set of attributes have been obtained.

$\{b\} \leftrightarrow E_2$, $\{b,c\} \leftrightarrow \{C,D,E\}$, $\{a,b,d\} \leftrightarrow \{A,B,C,E\}$, $\{a,b,c,d\} \leftrightarrow \{C,E\}$, $E_2 \leftrightarrow \varnothing$.

Having done the calculus for two different cases we have obtained the following conclusions.

The number of the homogeneous groups of the regions of Russian Federation represented by Moore's closures is much bigger than we have it for Ukraine. The first reason is that the data basis of Russia includes the input elements (social and economic indicators) that are more similar to each other while the economic indicators for the second country have significant dissimilarity between themselves. The second reason is that the method we have used to form the homogeneous groups is very sensible to any minimum changes that data basis have. We have observed this fact setting different levels (degrees) of the membership function for the groups of attributes.

We have obtained very detailed panorama of the economic situation for both countries. The bigger number of the homogeneous groups we have the easier and more complete economic analysis of the regions we can do, and, as a consequence, the more accurate information we contribute to the decision-making process at any required period of time.

Using this type of mathematical tools based on the theory of affinities we have obtained not only the exact number of the homogeneous groups of the economic regions but also the indicators (attributes) that the groups have in common. This information gives us an additional data basis to estimate the level of economic development of each region.

Galois lattices give us a complete structural ordering of the results obtained for two different cases providing by this manner a better visualization of an actual economic situation of all regions.

9. Conclusion

In this article, we proposed an alternative method for the grouping process based on the theory of uncertainty. With this study we pretended to demonstrate the importance and efficiency of the mathematical tools we used for implementing the process of groups (clusters) formation in the practice. The considered technique takes into account the increasing level of impreciseness of the economic information a company has to deal with in its everyday activity. It helps to reduce the complexity of the obtained economic data and set solid basis for the next step of the information procedure and also for the decision making process. This technique is very useful when the company has to deal with a great amount of economic data.

In the first part of the paper we gave brief explanations regarding the concepts such as grouping (clustering) process, similarity and affinity. We also gave the references to the authors who made a significant contribution to the subject of clustering process and fuzzy grouping process, in particular. We gave the theoretical explications of the following questions: 1) generalization of the notion of similarity by using the theory of affinities; 2) transition from one axiomatic to the other in pretopologies; 3) obtaining the strictest uncertain pretopology determined as uncertain pretopology of Moore; 4) process for obtaining maximum groupings in the cases of fuzzy rectangular relations by using a Moore's closure and Moore's closed set.

To demonstrate the method's efficiency we illustrated an empirical example to partition economic regions of Russian Federation and Ukraine into groups due to the different levels of homogeneity. The obtained results allow making a proper selection of one group of the regions or another according to the required levels of economic indicators. As a conclusive phase of our empirical example we introduced Galois lattices in order to form a clearer vision of a regional economic situation in the countries and to give an additional support tool to the decision-making process.

References

1. Brouwer, R.: Clustering feature vectors with mixed numerical and categorical attributes. International Journal of Computational Intelligence Systems, Vol. 1, No. 4, 285-298. (2008)
2. Jain, A. K.; Dubes, R. C.: Algorithms for Clustering Data. Upper Saddle River, NJ, Prentice Hall. (1988)
3. Klosgen, W.; Zytkow, J. M.: Knowledge discovery in databases terminology. In Advances in Knowledge Discovery and Data Mining. Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. The MIT Press, 573-592. (1996)
4. Bunyak, F.; et al.: Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets. Advances in Experimental Medicine and Biology, Vol. 696, 413-424. (2011)
5. Kyung, M. W.; et al. Breast tumor classification using fuzzy clustering for breast elastography. Ultrasound in Medicine and Biology, Vol. 37, No. 5, 700-708. (2011)
6. Folkesson, J.; et al.: Local bone enhancement fuzzy clustering for segmentation of MR trabecular bone images. Medical Physics, Vol. 37, No. 1, 295-302. (2010)
7. He, R.; et al.: Generalized fuzzy clustering for segmentation of multi-spectral magnetic resonance images. Computerized Medical Imaging and Graphics, Vol. 32, No. 5, 353-366. (2008)
8. Sakka, E.; et al.: Classification algorithms for microcalcifications in mammogram. Oncology Reports, Vol. 15, 1049-1055. (2006)
9. Hirano, S.; et al.: Comparison of clustering methods for clinical databases. Information Sciences, Vol. 159, 155-165. (2004)
10. Kang, X.; Xu, M.: Forecast about the disaster of Zoumaling tunnel discharge with fuzzy clustering analysis. In Proc. of International Conference on Electric Technology and Civil Engineering (ICETCE), 380-383. (2011)
11. Chang, T.; et al.: Dissolved gas analysis in transformer oil using dynamic tunneling fuzzy c-means algorithm. High Voltage Engineering, Vol. 35, No. 9, 2181-85. (2009)
12. Tari, L.; et al.: Fuzzy c-means clustering with prior biological knowledge. Journal of Biomedical Informatics, Vol. 42, No. 1, 74-81. (2009)
13. Zio, E.; Baraldi, P.: Evolutionary fuzzy clustering for the classification of transients in nuclear components. Progress of Nuclear Energy, Vol. 46, No. 3, 282–296. (2005)
14. Zio, E.; Gola, G.: Neuro-fuzzy patterns classification for fault diagnosis in nuclear components. Annals of Nuclear Energy, Vol. 33, No. 5, 415–426. (2006)
15. Wong, F.; et al.: Civil engineering including earthquake engineering. In Zimmermann, H.J. (ed.) Practical Applications of Fuzzy Technologies. Kluwer Academia Press, Boston, London, Dordrecht, 207-236. (1999)
16. Eom, K.: Fuzzy clustering approach in supervised sea-ice classification. Neurocomputing, Vol. 25, 149–166. (1999)
17. Zhuge, Y.; et al.: Parallel fuzzy connected image segmentation on GPU. Medical Physics, Vol. 38, No. 7, 4365-71. (2011)
18. Zhang, Y.; et al.: Image segmentation using PSO and PCM with Mahalanobis distance. Expert Systems with Applications, Vol. 38, No. 7, 9036-40. (2011)

19. Sowmya, B.; Sheela, R.: Colour image segmentation using fuzzy clustering techniques and competitive neural network. Applied Soft Computing, Vol. 11, No, 3, 3170-78. (2011)
20. Jablonowski, M.: Implications of fuzziness for the practical management of high-stakes risks. International Journal of Computational Intelligence Systems, Vol. 3, No. 1, 1-7. (2011)
21. Gil Aluja, J.; et al.: Using homogeneous groupings in portfolio management, Expert Systems with Applications, Vol. 38, No. 9, 10950-58. (2011)
22. Wang, Y.; Lee, H.: A clustering method to identify representative financial ratios, Information Sciences, Vol. 178, 1087–97.
23. Gil Aluja, J.; Gil Lafuente, A. M.: Algorithms for the treatment of complex economic phenomena (In Spanish). CEURA, Madrid, Spain. (2007)
24. Gil Aluja, J.: Introduction to the theory of uncertainty in business management (In Spanish). Milladoiro, Vigo. (2002)
25. Gil Lafuente, J.; Gil Aluja, J.: Algorithms for excellence. Keys for being successful in sport management (In Spanish). Milladoiro, Vigo. (2002).
26. Gil Aluja, J.: Pretopology in management of uncertainty. Inaugural speech (In Spanish). Public University, Leon. (2001)
27. Gil Lafuente, A. M.: New strategies for financial analysis of a business (In Spanish). Ariel, Barcelona. (2001)
28. Gil Lafuente, J.: Model for the homogeneous groping of the sales force. Proceedings of Congress M.S. China, pp. 332-335. (2001)
29. Gil Aluja, J.: Elements of a theory of decision in uncertainty. Kluwer Academic Publishers, Netherlands. (1999)
30. Gil Aluja, J.: Interactive management of human resources in uncertainty. Kluwer Academic Publishers, London, Dordrecht. (1998)
31. Gil Lafuente, J.: Marketing for the new millennium (In Spanish). Pirámide, Madrid. (1997)
32. Gil Aluja, J.: The multicriteria selection of investments by means of the Galois lattice (In Spanish). Civita, Madrid, cap. 7, pp 139-157. (1995)
33. Kaufmann, A.; Gil Aluja, J.: Selection of affinities by means of fuzzy relation and Galois Lattices. Acts of the XI European Congress O.R., Aachen. (1991)
34. Klir, J.; Folger, A: Fuzzy sets, uncertainty and information. Prentice-Hall International, USA. (1988)
35. Kaufmann, A.: Méthodes et modeles de la recherche opérationnelle. Dunod, París, Vol 1, 65-72. (1970)
36. Kaufmann, A.; et al.: Cours de calcul booléien appliqué. Albin Michel, Francia. (1963)
37. Anderberg, M. R.: Cluster Analysis for Applications. Academic Press. (1973)
38. Everitt, B.: Cluster analysis. Heinemann Educational Books, 117-129. (1996)
39. Kaufman, L.; Rousseeuw, P.: Finding groups in data: an introduction to cluster analysis. John Wiley and Sons, New York. (1990)
40. Michalski, R. S.; Stepp, R. E.: Automated construction of classifications: conceptual clustering versus numerical taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5, 396-410. (1983)
41. Fisher, D. H.: Knowledge acquisition via incremental conceptual clustering. Machine Learning, Vol. 2, No. 2, 139-172. (1987)
42. Zadeh, L.: Fuzzy algorithms. Information and Control, Vol. 12, 94-102. (1968)
43. Li, G.; et al.: Clustering with instance and attribute level side information. International Journal of Computational Intelligence Systems, Vol. 3, No. 6, 770-785. (2010)
44. Kumar, P.: Fuzzy classifier design using modified genetic algorithm. International Journal of Computational Intelligence Systems, Vol. 3, No. 3, 334-342. (2010)

45. Butenko, S.; Gil-Lafuente, J.; Pardalos, P.: Optimal strategies in sports economics and management. Springer-Verlag Berlin Heidelberg, pp. 1-15. (2010)
46. Brouwer, R.: Fuzzy clustering of feature vectors with some ordinal valued attributes using gradient descent for learning. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 6, 195-218. (2008)
47. Honda, K.; Ichihashi, H.: Regularized linear fuzzy clustering and models. IEEE Transactions on Fuzzy Systems, Vol. 13, No. 4, 508–516. (2005)
48. Yang, M.; et al.: Fuzzy clustering algorithms for mixed feature variables, Fuzzy Sets and Systems, Vol. 141, 301-317. (2004)
49. Miyamoto, S.: Information clustering based on fuzzy multisets. Information Processing and Management, Vol. 39, 195-213. (2003)
50. Hoppner, F.; et al.: Fuzzy Cluster Analysis. John Wiley and Sons, New York. (1999)
51. Kaufmann, A.: Introduction to the theory of fuzzy sets for use by engineers, Vol. 4, Additions and new applications (In French). Masson, Paris. (1977)
52. Zadeh, L.: Similarity relations and fuzzy orderings. Information Sciences, Vol. 3, No. 2, 177-200. (1971)
53. Ruspini, E.: Numerical methods for fuzzy clustering. Information Sciences, Vol. 2, 319-350. (1970)
54. Zadeh, L.: Fuzzy algorithms. Information and Control, Vol. 12, 94-102. (1968)
55. Dolecki, S.; et al.: Weak regularity and consecutive topologizations and regularizations of pretopologies. Topology and its Applications, Vol. 156, No. 7, 1306-14. (2009)
56. Höhle, U.: Many valued topology and its applications. Kluwer Academic, Dordrecht. (2001)
57. Höhle, U.; et al.: A general theory of fuzzy topological spaces. Fuzzy Sets and Systems, Vol. 73, 131-149. (1995)
58. Kaufmann, A.: Ordinary pretopology and fuzzy pretopology (In French). Working Note 115. (1983)
59. Kaufmann, A.: Introducción a la combinatoria y sus aplicaciones. Campania Editorial Continental, Barcelona. In Spanish. (1971)
60. Gil Aluja, J.: Clan, affinities and Moore’s fuzzy pretopology. Fuzzy Economic Review, Vol. VIII, No. 2, 3-24. (2003)
61. Gil Aluja, J., Gil Lafuente, A. M., Mérigo, J. M.: Using homogeneous groupings in portfolio management. International Journal of Expert Systems with Applications, Vol. 38, No. 9, 10950-58. (2011)
62. Mérigo, J. M., Gil Lafuente, A. M.: Decision-making techniques with similarity measures and OWA operators. Statistics & Operations Research Transactions, Vol. 36, No. 1, 81-102. (2012)
63. Gil Lafuente, A. M., Barcellos Paula, L.: Approaches to managing sustainability among enterprises. Decision making systems in business administration. Vol. 8, pp. 27 -36. (2012)
64. Keropyan, A., Gil Lafuente, A. M.: A personal selection model using Galois group theory, Kybernetes, Vol. 42, No. 5, 711-719. (2013)
65. Gil Lafuente, J.: Are there clans in a wardrobe? “Fuzzy-Pichat” will find out it! Proceedings of the National Congress. Vol. 1, pp. 90. (2008)
66. Gil Lafuente, A. M.; Gil Lafuente, J.: Models and algorithms for the treatment of creativity in business management (In Spanish). Milladoiro, Vigo. (2007)
67. Gil-Aluja, J.; Gil-Lafuente, A. M.; Klimova, A.: M-attributes algorithm for the selection of a company to be affected by a public offering. International Journal of Uncertainty, Fuzziness Knowledge-Based Systems, Vol. 17, No. 3, 333-343. (2009)
68. Sapena, A.: A contribution to the study of fuzzy metric spaces. Applied General Topology, Vol. 2, No. 1, 63-76. (2001)

Anna M. Gil-Lafuente is an Associate Professor in the Department of Business Administration at the University of Barcelona, Spain. She received a MSc and a PhD degree in Business Administration from the University of Barcelona. She has published more than 150 papers in journals, books and conference proceedings including journals such as *Information Sciences*, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* and *Expert Systems with Applications*. She is co-editor-in-chief of 8 journals published by the Association for Modeling and Simulation in Enterprises (AMSE). She is currently interested in Decision Making, Uncertainty and Finance.

Anna Klimova is a Professor in the Department of Business Administration at the University of Barcelona, Spain. She received a PhD degree in Business Administration from the University of Barcelona. She has published more than 20 papers in journals and conference proceedings. She is currently interested in Decision Making, Uncertainty and Finance.

Received: January 25, 2013; Accepted: December 24, 2014