Embedding Tolerance Relations in Formal Concept Analysis: An Application in Information Fusion[*]

Mehdi Kaytoue   Zainab Assaghir   Amedeo Napoli
Inria Nancy Grand Est / Loria – Campus Scientifique
54506 Vandœuvre-lès-Nancy, France
{kaytouem,assaghiz,napoli}@loria.fr

Sergei O. Kuznetsov
State University Higher School of Economics
Pokrovskiy Bd. 11 – 101000 Moscow, Russia
skuznetsov@hse.ru

ABSTRACT

This paper shows how to embed a similarity relation between complex descriptions in concept lattices. We formalize similarity by a tolerance relation: objects are grouped within a same concept when having similar descriptions, extending the ability of FCA to deal with complex data. We propose two different approaches. A first classical manner defines a discretization procedure. A second way consists in representing data by pattern structures, from which a pattern concept lattice can be constructed directly. In this case, considering a tolerance relation can be mathematically defined by a projection in a meet-semi-lattice. This allows to use concept lattices for their knowledge representation and reasoning abilities without transforming data. We show finally that resulting lattices are useful for solving information fusion problems.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods; I.5.3 [**Clustering**]: Similarity measures

General Terms

Theory, Management, Experimentation

1. INTRODUCTION

In this paper, we are interested in analyzing complex real-world data. Data can be analyzed with knowledge discovery methods s.t. Formal Concept Analysis (FCA) [3], a mathematically well founded classification framework allowing to derive implicit relationships from a set of objects and their attributes. The main structure which is built is a concept lattice, that can be represented by a diagram where classes of objects and ordering relations between classes can be drawn and interpreted and used for knowledge management, data

---

mining, etc. To handle complex data with FCA one requires scaling (discretization), which transforms many-valued attributes to binary attributes. A scaling is a way of data interpretation, different scaling may result in different concept lattices.

We propose a classification approach based on FCA and an appropriate scaling that can be applied to real-world objects described by numerical attributes and taking into account similarity between attribute values. The mathematical formalization of similarity relies on a *tolerance relation* which is reflexive and symmetric. A tolerance relation can be used for building tolerance classes of similar objects that are reused to properly define a scaling for initial numerical data allowing FCA to be applied. Besides scaling, it is possible to directly process complex data using the so-called pattern structure approach. This extension of FCA is recalled here and applied to intervals. Contrasting the large body of work on discretization for numerical classification methods [8], this paper introduces and explains relationships between standard FCA, scaling, pattern structures, and similarity.

The paper is organized as follows. Section 2 presents preliminaries on FCA and scaling procedures. Section 3 introduces tolerance relations, the way how scales can be designed from numerical data, and how numerical concept lattices are built. Section 4 describes pattern structures in FCA and the use of tolerance relations in this framework. Section 5 describes a real-world experiment in agronomy, where the resulting lattice materializes information fusion and can be rather easily interpreted by agronomy experts.

Table 1: Numerical data. Table 2: Formal context.

	m_1	m_2	m_3		m_1	m_2	m_3
g_1	6	0	$[1,2]$	g_1		$\times$	$\times$
g_2	8	4	$[2,5]$	g_2		$\times$	$\times$
g_3	11	8	$[4,5]$	g_3	$\times$		$\times$
g_4	16	8	$[6,9]$	g_4	$\times$		
g_5	17	12	$[7,10]$	g_5	$\times$		

2. FORMAL CONCEPT ANALYSIS (FCA)

FCA starts with a *formal context*, which is a triple (G, M, I), where G denotes a set of objects, M a set of attributes, and $I \subseteq G \times M$ a binary relation between G and M. The statement $(g, m) \in I$ is interpreted as "the object g has attribute m" (Table 2). The two derivation operators $(\cdot)'$ define a Galois connection between the powersets $(2^G, \subseteq)$ and $(2^M, \subseteq)$. $A' = \{m \in M \mid \forall g \in A : gIm\}$, for $A \subseteq G$, and $B' = \{g \in G \mid \forall m \in B : gIm\}$ for $B \subseteq M$. For $A \subseteq G$, $B \subseteq M$, a pair (A, B), such that $A' = B$ and $B' = A$, is called a *(formal) concept*, e.g. $(\{g_3, g_4, g_5\}, \{m_1\})$. In
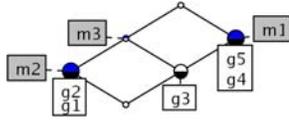
Figure 1: Concept lattice associated with Table 2.

(A, B), the set A is called the *extent* and the set B the *intent* of the concept (A, B). Concepts are partially ordered by $(A_1, B_1) \le (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \ (\Leftrightarrow B_2 \subseteq B_1)$, e.g. the concept $(\{g_3\}, \{m_1, m_3\})$ is a sub-concept of $(\{g_3, g_4, g_5\}, \{m_1\})$. With respect to this partial order, the set of all formal concepts forms a complete lattice called the *concept lattice* of the formal context (G, M, I). Figure 1 shows the concept lattice associated with Table 2. On the diagram, each node denotes a concept while a line denotes an order relation between two concepts. Due to *reduced labeling*, the extent of a concept has to be considered as composed of all objects lying in the extents of its sub-concepts. Dually, the intent of a concept is composed of all attributes in the intents of its super-concepts.

Non-binary data (Table 1) are described by a many-valued context (G, M, W, I), where W denotes a set of attribute values, such that $(g, m, w) \in I$. The fact that attribute m takes value w for object g is denoted by $m(g) = w$. A standard approach in FCA for constructing a concept lattice from a many-valued context (G, M, W, I) consists in applying conceptual scaling, where a scale for a many-valued attribute is given by a transformation of attribute-value pairs into a set of binary attributes. For example, Table 1 can be transformed into Table 2 where the scale for m_1 is given by $\{m_1 \ge 10\}$, resp. $\{m_2 \le 6\}$ and $\{m_3 \le 5\}$ for m_2 and m_3. The choice of a scale is arbitrary and usually leads to loss of information (links and closeness between values) and border problems, but it remains an important technique for binarizing complex data [3, 8].

3. TOLERANCE-BASED SCALING

Introduction and definitions. Similarity has been studied from many points of view in artificial intelligence and pattern recognition [7]. For example, considering documents described by their attributes, e.g. keywords, similarity of documents x and y can be defined by non-emptiness of the set of their common attributes, $x' \cap y' \ne \emptyset$. The similarity is reflexive and symmetric, but not necessarily transitive. Following this idea, a tolerance relation captures the characteristics of a similarity [6].

DEFINITION 3.1. *For a set G, a binary relation $T \subseteq G \times G$ is called* tolerance *if:*

    (i) $\forall x \in G \ xTx$ (reflexivity)
    (ii) $\forall x, y \in G \ xTy \rightarrow yTx$ (symmetry)

Let us consider now a set of objects G, a tolerance relation T, and a formal context (G, G, T). First, some objects, say g_1 and g_2, are observed to be similar, i.e. $g_1 T g_2$. Then pairs of the tolerance relation lead to a class of similar objects or "class of similarity". Moreover, among the classes of similarity, some classes are maximal meaning that the class is not included in any larger class.

DEFINITION 3.2. *Given a set G, a subset $K \subseteq G$, and a tolerance relation T on G, K is a class of tolerance if:*

    (i) $\forall x, y \in K \ xTy$ (pairwise similarity)
    (ii) $\forall z \notin K, \exists u \in K \ \neg(zTu)$ (maximality)

An arbitrary subset of a class of tolerance is a preclass.

Now, let us consider the classes of tolerance associated with the formal context (G, G, T). The class of tolerance of an object g has to be considered along two dimensions: *(i)* the class is defined as the set of all objects which are tolerant with g, *(ii)* the class is maximal in the sense that objects in the class are pairwise similar, and adding any other object in the class results in some pairs of non tolerant objects. A class of tolerance may be given a name which can be further used as an "attribute name" that describes the object. The result is a formal context (G, M, I) where I associates any object in G with its classes of tolerance $m \in M$.

Let us return to objects and numerical attributes of Table 1. Intuitively, two objects g_1 and g_2 are similar for a set of attributes if the values for each attribute are "similar". Similarity (or closeness) of two numerical values can be measured by the difference of these two values: $|m_1(g_1) - m_1(g_2)|$. Then, two numerical values are similar when their difference is lower than a *similarity threshold* θ expressing the maximal variation allowed between two similar values. More precisely, given two numbers $a, b \in \mathbb{R}$ and a similarity threshold θ, a similarity relation $\simeq_\theta$ is defined as:
$$a \simeq_\theta b \Leftrightarrow |a - b| \le \theta$$
This similarity relation $\simeq_\theta$ is reflexive and symmetric but not necessarily transitive, e.g. $1 \simeq_2 3$, $3 \simeq_2 5$, but $1 \not\simeq_2 5$.

The same holds when attributes are interval-valued by defining a similarity $\simeq_\theta$ as follows, with $a, b, c, d \in \mathbb{R}$:
$$[a, b] \simeq_\theta [c, d] \Leftrightarrow max(b, d) - min(a, c) \le \theta$$
provided that $|a - b| \le \theta$ and that $|c - d| \le \theta$. In other words, two numerical intervals are θ-*similar* if the length of their "convex hull" is not larger than a threshold θ.

Classes of tolerance for numerical attributes. Let us consider a numerical many-valued context (G, M, W, I) where the range W_m of an attribute m is such that $W_m \subseteq W \subset \mathbb{R}$. Each attribute has a different range and different similarities and thresholds θ have to be defined. However, data can be normalized leading to a single threshold.

Given an attribute $m \in M$, let us consider the formal context $(W_m, W_m, \simeq_\theta)$. Similar objects in W_m are related w.r.t. $\simeq_\theta$. For example, given $\theta = 5$ and m_1 in Table 1, the formal context $(W_{m_1}, W_{m_1}, \simeq_5)$ can be read in Figure 2 (left). As $\simeq_5$ is symmetric and reflexive, $(W_{m_1}, W_{m_1}, \simeq_5)$ contains a diagonal of crosses and the associated concept lattice (see Figure 2 (right)) is also symmetric.

PROPOSITION 3.1. *Given a context $(W_m, W_m, \simeq_\theta)$ and the associated lattice, any concept (A, B) is such that either $A \subset B$, $B \subset A$, or $A = B$. For each concept (A, B), the pair (B, A) is also a concept.*

For example, the upper right concept on Figure 2 (right) can be read as $(\{8, 6, 11, 16\}, \{11\})$ and has a corresponding concept $(\{11\}, \{8, 6, 11, 16\})$ lower still on the right. One consequence of the above proposition is that the concept lattice can be separated in two parts w.r.t. the mapping $(A, B) \mapsto (B, A)$. In [3], such a mapping is called a *polarity*, i.e. an order-reversing bijection inverse of itself, and the resulting concept lattice is a *polarity lattice*. Then, the set of all concepts (A, B) such that $A = B$ forms an *axis of polarity* of the concept lattice.

Concepts $(\{6, 8, 11\}, \{6, 8, 11\})$, $(\{16, 17\}, \{16, 17\})$ and $(\{11, 16\}, \{11, 16\})$, form the axis of polarity of the lattice on Figure 2 (right). The set of all concepts (C, D) such that $(A, B) \le (C, D)$, denoted by U, forms the upper part of the concept lattice. Dually, the set of all concepts (E, F) such that $(E, F) \le (A, B)$, denoted by L, forms the lower part of the concept lattice. If $(A, B) \in U$ then $(B, A) \in L$ and $B \subset A$. Dually, if $(A, B) \in L$ then $(B, A) \in U$ and $A \subset B$.

m_1	6	8	11	16	17
6	×	×	×		
8	×	×	×		
11	×	×	×	×	
16			×	×	×
17				×	×
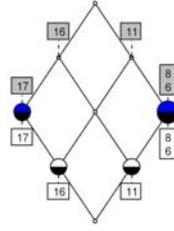
Figure 2: Symmetric context formalizing a tolerance relation (left) and its concept lattice (right).

PROPOSITION 3.2. *Let (A, B) be a concept of the axis of polarity, i.e. $A = B$. Then, A (or B) is a set of maximal pairwise similar values, i.e. A determines a class of tolerance. Let (C, D) a concept in U but not in the axis of polarity, i.e. $D \subset C$. D is a preclass of tolerance and C is the set of all values similar to values in D.*

The intents of the concepts in the upper part of the lattice –or dually the extents in the lower part– are partially ordered and determine sets of similar values. Among these intents, the intents in the axis of polarity are maximal and are classes of tolerance, and the other intents are only preclasses of tolerance. For example, taking $\theta = 5$ and m_1 in Table 1, there are 5 intents, namely $\{16\}$, $\{11\}$, $\{16, 17\}$, $\{11, 16\}$, and $\{6, 8, 11\}$, where the three last intents are tolerance classes. When there is no ambiguity, we use the term of "class of similarity" for a class or a "preclass of tolerance".

We have made precise how a partially ordered set of classes of similarity can be built from a many-valued context. Now, classes of similarity have to be named before being used as attribute names for scaling the original many-valued context. Actually, the name of the elements of the scale can be related to the semantic content of the corresponding class of similarity and to the name of the original attribute that is scaled. Here, an element of the scale is named by a pair associating the name of the original attribute and either the content of the class of similarity or their convex hull, e.g $\{16, 17\}$ or $[16, 17]$ for m_1.

Table 3: Formal context obtained from Table 1 handling classes of tolerance of attributes m_1 and m_2.

	$(m_1, 11)$	$(m_1, 16)$	$(m_1, [6, 11])$	$(m_1, [11, 16])$	$(m_1, [16, 17])$	$(m_2, 4)$	$(m_2, 8)$	$(m_2, [0, 4])$	$(m_2, [4, 8])$	$(m_2, [8, 12])$
g_1			×					×		
g_2			×			×		×	×	
g_3	×		×	×			×		×	×
g_4		×		×	×		×		×	×
g_5					×					×

Building a "numerical concept lattice". Consider the numerical context (G, W, M, I) in Table 1. Three sets of classes of similarity (one for each attribute) are computed thanks to three tolerance relations relying on three different similarities $\simeq_\theta$, and extracted from the symmetric concept lattices associated with each tolerance relation. The transformation of the original context (G, W, M, I) into the derived context (G, N, J) with G is the set of original objects, $N = \bigcup_{m \in M} (\{m\} \times C_m)$ with C_m is the set of all classes of similarity of attribute m, and $(g, (m, C_m)) \in J$ means that the value of object g in the many-valued context, i.e. $m(g)$, belongs to class C_m. The derived binary context associated with Table 1 is given in Table 3 for attributes m_1 and m_2 where the thresholds are $\theta = 5$ for m_1 and $\theta = 4$ for m_2 (and $\theta = 5$ for m_3).

4. TOLERANCE IN PATTERN STRUCTURES

The preceding work takes place in standard FCA with scaling. An extension consists on working directly on complex data, without scaling. For this, a *pattern structure* is a generalization of a formal context for complex data [2]. First a similarity operation $\sqcap$ is defined on object descriptions. This operation is idempotent, commutative and associative, which allows for defining a natural partial order on D.

Formally, let G be a set of objects, $(D, \sqcap)$ be a meet-semilattice of object descriptions and $\delta : G \to D$ a mapping associating an object with its description. Then $(G, (D, \sqcap), \delta)$ is called a *pattern structure*. Elements of D are called *patterns* and are ordered by the subsumption relation $\sqsubseteq$: given $c, d \in D$, $c \sqsubseteq d \Leftrightarrow c \sqcap d = c$. A pattern structure $(G, (D, \sqcap), \delta)$ gives rise to the following derivation operators $(\cdot)^\square$, given $A \subseteq G$ and $d \in (D, \sqcap)$: $A^\square = \prod_{g \in A} \delta(g)$ and $d^\square = \{g \in G | d \sqsubseteq \delta(g)\}$. These operators form a Galois connection between $(2^G, \subseteq)$ and $(D, \sqsubseteq)$. *(Pattern) concepts* of $(G, (D, \sqcap), \delta)$ are pairs of the form (A, d), $A \subseteq G$, $d \in (D, \sqcap)$, s.t. $A^\square = d$ and $A = d^\square$. For a pattern concept (A, d), d is called *intent* and is a description of all objects in A, called *extent*. Partially ordered by $(A_1, d_1) \leq (A_2, d_2) \Leftrightarrow A_1 \subseteq A_2$ ($\Leftrightarrow d_2 \sqsubseteq d_1$), the set of all concepts forms a so-called *pattern concept lattice*.

Interval pattern structures [4]. A numerical dataset is represented by a many-valued context (G, M, W, I). Objects are described by numbers or intervals, each one standing for a given attribute, and hence *interval vectors* are introduced as patterns. When c and d are interval vectors, we write $c = \langle [a_i, b_i] \rangle_{i \in [1, |M|]}$ and $d = \langle [c_i, d_i] \rangle_{i \in [1, |M|]}$. Interval vectors may be partially ordered within a meet-semi-lattice as follows. Given two interval vectors $c = \langle [a_i, b_i] \rangle_{i \in \{1, ..., |M|\}}$, and $d = \langle [c_i, d_i] \rangle_{i \in \{1, ..., |M|\}}$, $c \sqcap d = \langle [min(a_i, c_i), max(b_i, d_i)] \rangle_{i \in \{1, ..., |M|\}}$ meaning that a "convexification" of intervals on each vector dimension is operated. The meet operator induces the following subsumption relation $\sqsubseteq$ on interval patterns $\langle [a_i, b_i] \rangle \sqsubseteq \langle [c_i, d_i] \rangle \Leftrightarrow [a_i, b_i] \supseteq [c_i, d_i]$, $\forall i \in \{1, ..., |M|\}$. In Table 1, description of g_1 is $\delta(g_1) = \langle [6, 6], [0, 0], [1, 2] \rangle$. We have $\delta(g_1) \sqcap \delta(g_2) = \langle [6, 8], [0, 4], [1, 5] \rangle$, and therefore $\langle [6, 8], [0, 4], [1, 5] \rangle \sqsubseteq \langle [6, 6], [0, 0], [1, 2] \rangle$. The Galois connection is illustrated as follows. $\{g_1, g_3\}^\square = \langle [6, 11], [0, 8], [1, 5] \rangle$ and $\langle [6, 11], [0, 8], [1, 5] \rangle^\square = \{g_1, g_2, g_3\}$, making the pair $(\{g_1, g_2, g_3\}, \langle [6, 11], [0, 8], [1, 5] \rangle)$ a pattern concept.

Tolerance relation in pattern structures. Given $a, b, c, d \in \mathbb{R}$, a parameter $\theta \in \mathbb{R}$, we say that two descriptions (intervals) $[a, b]$ and $[c, d]$ are similar if: $max(b, d) - min(a, c) \leq \theta$, i.e. size of $[a, b] \sqcap [c, d]$ does not exceed θ.

Then, though each interval from a semi-lattice $(D, \sqcap)$ describes a preclass of tolerance, some intervals may not be "maximal", i.e. not describing classes of tolerance. Below, we show how to replace any interval by its "maximal" interval with a *projection* in a meet-semi-lattice.

First, consider the meet-semi-lattice $(D, \sqcap)$ of interval values for a given attribute. Then, for any interval $x \in D$, we define the ball $B(x, \theta)$ as the set of intervals in D similar to x as follows: $B(x, \theta) = \{y \in D \mid y \simeq_\theta x\}$. This ball with center x and diameter θ contains all intervals y such that x and y are *similar*.

Now, we remove from $B(x, \theta)$ any pair of intervals that are not pairwise similar, and build an interval with the left border (resp. right border) as the minimum (resp. maximum) of all intervals. This can be done by replacing any x of the

meet-semi-lattice of intervals by the meet of all intervals y from the ball $B(x,\theta)$ that are not dissimilar with another element y' of this ball:

$$\psi(x,\theta) := \prod\{y \in B(x,\theta) | \forall y'[y' \in B(x,\theta) \rightarrow y \simeq_\theta y']\}$$

ψ maps any $x \in D$ to a representation of its class of tolerance, i.e. the associated maximal set of pairwise similar elements. For example, with attribute m_3, we have $\psi([2,5],5) = [1,5]$ and $[1,5]$ is the similarity (in the sense of operation $\sqcap$) of all elements of the class of tolerance containing $[2,5]$, i.e. the class $\{[1,2],[2,5],[4,5]\}$. When the size of an interval exceeds θ, the ball of similar patterns is empty. ψ is a mapping that takes any $x \in D$ to an element $\psi(x) \in (D,\sqcap)$ such that $\psi(x) \sqsubseteq x$, meaning that ψ is contractive. In sense of [2], ψ is a *projection* in the semi-lattice $(D,\sqcap)$ as also monotone and idempotent. Moreover, any projection of a complete semi-lattice $(D,\sqcap)$ is $\sqcap$-preserving, i.e. for any $x,y \in D$, $\psi(x\sqcap y) = \psi(x)\sqcap\psi(y)$ [2]. Thereby, the projection may be computed in advance, replacing each pattern by a "weaker" or "more general" pattern. It comes with a loss of information, e.g in previous example $[2,5]$ replaced by $[1,5]$ which is more imprecise. However, this loss of information is controlled by θ: the projected pattern structure preserves the similarity between descriptions in the original pattern structure, and keeps the same representation formalism while embedding a tolerance relation.

5. AN INFORMATION FUSION PROBLEM

Agronomists compute indicators for evaluating the impact of agricultural practices on the environment. The risk level for a pesticide to reach groundwater is computed by the indicator I_{gro} in [1]. Based on the value of I_{gro}, agronomists try to make a diagnosis of agronomic know-how w.r.t. the use of pesticides. Pesticide characteristics depend on the chemical characteristics of the product while pesticide period application and field characteristics depend on domain knowledge. This knowledge lies in information sources such as books, databases, and expert knowledge in agronomy. Moreover, values for some characteristics may vary w.r.t. information sources. Here, we are interested in the analysis of practices through the use of *glyphosate* in different countries w.r.t. farmers habits. Below, three characteristics of glyphosate, namely $DT50$, koc, and ADI (detailed in [5]), are given in Table 4 (simplified data), according to 12 different information sources, but the latter are not always in agreement. Then, it can be interesting for experts in agronomy to analyse such a table from the point of view of information fusion: which are the sources being in agreement and at which level? We apply our framework on similarity and scaling to build a concept lattice from Table 4. Three thresholds are defined according to experts [5]: $\theta = 100$ for $DT50$, $\theta = 2200$ for koc, and $\theta = 0$ for ADI. The resulting lattice (Figure 3) shows an interesting classification of information sources w.r.t. information fusion. Each concept in the lattice is composed of an extent with a maximal set of sources in agreement w.r.t. the interval of values in the intent. According to experts in agronomy, the lattice on Figure 3 is a good witness of the diversity of practices and of the agreement degree between sources (see [5]).

6. CONCLUSION

This paper shows how to build concept lattices from complex data, where concepts are sets of similar objects, (i) with

Table 4: Characteristics of pesticide *glyphosate*.

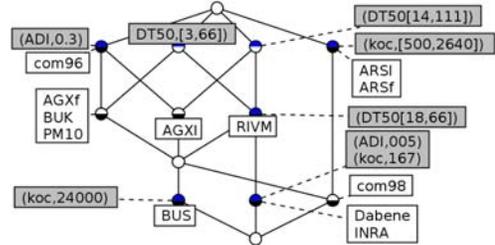|	$DT50$ (days)	koc (L/kg)	ADI (g/kg/day)
BUS	47	24000	0.3
PM10	[3,60]	[25,68000]	0.3
INRA	[38,60]	167	0.05
Dabene	[38,60]	167	0.05
ARSf	[2,174]	[500,2640]	[0.05,0.3]
ARSl	[2,174]	[500,2640]	[0.05,0.3]
Com96	[2,174]	[25,68000]	0.3
Com98	[38,60]	[500,2640]	0.3
RIVM	[18,66]	[3566,40420]	[0.05,0.3]
BUK	[3,60]	[25,68000]	0.3
AGXf	[8,30]	[301,59000]	0.3
AGXl	[14,111]	[301,59000]	0.3

Figure 3: Concept lattice raised from Table 4.

an appropriate scaling and classical FCA, (ii) by projecting a meet-semi-lattice of patterns. (ii) allows to use FCA for knowledge representation and reasoning without transforming data and can be generalised to any structured data for which a similarity measure can be defined. Finally, we show that the resulting lattice can characterize subsets of sources with similar and precise information for fusion issues. It remains to carry out a deep analysis on links between discretization methods and projections of semi-lattices, while an interesting perspective concerns association rules, e.g. $(ADI = 0.05 \wedge koc = 167) \rightarrow (DT50 \in [18,66])$ with confidence 100% can be read in Figure 3.

7. REFERENCES

[1] C. Bockstaller, P. Girardin, and H. Van der Werf. Use of agro-ecological indicators for the evaluation of farming systems. *European journal of agronomy*, 7:261–270, 1997.

[2] B. Ganter and S. O. Kuznetsov. Pattern structures and their projections. In *Int. Conf. on Conceptual Structures*, LNCS 2120, pages 129–142, 2001.

[3] B. Ganter and R. Wille. *Formal Concept Analysis*. Springer, mathematical foundations edition, 1999.

[4] M. Kaytoue, S. Duplessis, S. Kuznetsov, and A. Napoli. Two FCA-Based Methods for Mining Gene Expression Data. In *ICFCA*, LNCS 5548, pages 251–266, 2009.

[5] M. Kaytoue, S. O. Kuznetsov, Z. Assaghir, and A. Napoli. Embedding Tolerance Relations in Concept Lattices - An application in Information Fusion. Research Report RR-7353, INRIA, 2010.

[6] S. Kuznetsov. Galois Connections in Data Analysis: Contributions from the Soviet Era and Modern Russian Research. In *ICFCA*, LNCS 3626, pages 196–225, 2005.

[7] A. Tversky. Features of Similarity. *Psychological Review*, 84(4):327–352, 1977.

[8] Y. Yang, G. I. Webb, and X. Wu. Discretization methods. In *The Data Mining and Knowledge Discovery Handbook*, pages 113–130. Springer, 2005.