Managing Information Fusion with Formal Concept Analysis

Zainab Assaghir[1], Mehdi Kaytoue[1], Amedeo Napoli[1], and Henri Prade[2]

[1] Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)
Campus Scientifique, B.P. 70239 – 54500 Vandœuvre-lès-Nancy – France
[2] Institut de Recherche en Informatique de Toulouse (IRIT)
118 Route de Narbonne – 31062 Toulouse – France
assaghiz@loria.fr, kaytouem@loria.fr, napoli@loria.fr, prade@irit.fr

Abstract. The main problem addressed in this paper is the merging of numerical information provided by several sources (databases, experts...). Merging pieces of information into an interpretable and useful format is a tricky task even when an information fusion method is chosen. Fusion results may not be in suitable form for being used in decision analysis. This is generally due to the fact that information sources are heterogeneous and provide inconsistent information, which may lead to imprecise results. In this paper, we propose the use of Formal Concept Analysis and more specifically pattern structures for organizing the results of fusion methods. This allows us to associate any subset of sources with its information fusion result. Once a fusion operator is chosen, a concept lattice is built. With examples throughout this paper, we show that this concept lattice gives an interesting classification of fusion results. When the fusion global result is too imprecise, the method enables the users to identify what maximal subset of sources that would support a more precise and useful result. Instead of providing a unique fusion result, the method yields a structured view of partial results labelled by subsets of sources. Finally, an experiment on a real-world application has been carried out for decision aid in agricultural practices.

1 Introduction

In this paper, we present a method for managing information fusion based on Formal Concept Analysis (FCA) when information is numerical. The problem of information fusion is encountered in various fields of application, e.g sensor fusion, multiple source interrogation systems. Information fusion consists of merging, or exploiting conjointly, several sources of information for answering questions of interest and make proper decisions [1]. A fusion operator is an operation summarizing all information given by sources into an interpretable information, for example the interval intersection for numerical information.

Several fusion operators were proposed for combining uncertain information [2, 3, 4, 5, 6, 7] and no universal method is available [2]. Dubois and Prade [2] overviewed how fuzzy set theory can address the information fusion problem

V. Torra, Y. Narukawa, and M. Daumas (Eds.): MDAI 2010, LNAI 6408, pp. 104–115, 2010.
© Springer-Verlag Berlin Heidelberg 2010

Table 1. Information dataset given by sources

	m_1	m_2
g_1	$[1, 5]$	$[1, 9]$
g_2	$[2, 3]$	$[1, 3]$
g_3	$[4, 7]$	$[6, 7]$
g_4	$[6, 10]$	$[8, 9]$

Table 2. A formal context

	m_1	m_2
g_1	×	
g_2	×	
g_3	×	×
g_4		×

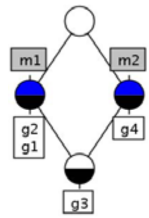

Fig. 1. Concept lattice raised from Table 2

and proposed several fusion operators for numerical information. More recently, a fusion operator based on the notion of Maximal Consistent Subset (MCS) has been proposed for finding a global point of view when no meta-knowledge is available about sources (reliability, conflict) [8, 9]. These works apply the fusion operator on the set of all sources and consider the resulting information. Other approaches define their proper fusion operator in a lattice structure to combine symbolic information [6, 7].

In this work, we use FCA to study all subsets of sources and their information fusion results. The main ability of FCA is to produce formal concepts corresponding to maximal sets of sources associated with a fused information. The concepts are ordered and form a structure called concept lattice. We show that this lattice contains the information fusion result considering all sources proposed by [2, 8, 9]. Moreover, the lattice is meaningful for organizing information fusion results of different subsets of sources and allows more flexibility for the user. Moreover, the lattice keeps a track of the origin of the information such as presented in [3] for the fusion of symbolic information.

This work can be used in many applications where it is necessary to find a suitable value summarizing several values coming from multiple sources. Here, we use an experiment in agronomy for decision helping in agricultural practices.

The paper is organized as follows. Section 2 presents and illustrates the basics of fusion operators. Section 3 introduces the preliminaries on FCA and its generalization for handling numerical data. Then, Section 4 shows how FCA is well suited for organizing different information fusion results. Section 5 describes a real-world experiment: a concept lattice embedding fusion results is interpreted for making decisions about agricultural practices.

2 Basics of Numerical Information Fusion Operators

According to previous works, there are three kinds of behaviors for the fusion operators: conjunctive, disjunctive and trade-off operators [1, 2, 4].

Before introducing these operators, we introduce the following notations: n is the number of sources. $\mathbb{I}_m$ is the set of all values given for the variable m. $f_m : \mathbb{I}_m \rightarrow \mathbb{R}$ denotes a fusion operator returning the fusion result for variable m.

The *conjunctive operator* is the counterpart to a set intersection. The imprecision and the uncertainty in the information associated with the result of a conjunction is less than the imprecision or the uncertainty of each source alone. A conjunctive operator makes the assumption that all the sources are reliable, and usually results in a precise information. If there is some conflict in the information (i.e. at least one source is not fully reliable), then the result of the conjunction can be insufficiently reliable, or even empty. The conjunctive operator for a variable m is defined by $f_m(\mathbb{I}_m) = \bigcap_{i=1,\dots,n} I_i$, e.g., in Table 1, $f_{m_1}(I_1,\dots,I_4) = \emptyset$ represents the intersection of intervals of m_1 with $I_1 = [1,5], I_2 = [2,3], I_3 = [4,7]$ and $I_4 = [6,10]$.

The *disjunctive operator* is the counterpart to a set union. The uncertainty (or the imprecision) resulting from a disjunction is higher than the uncertainty (or the imprecision) of all sources together. A disjunctive operator makes the assumption that at least one source is reliable. The result of a disjunctive operator can be considered as reliable, but is also often (too) weakly informative. The disjunctive operator for the variable m, is defined by $f_m(\mathbb{I}) = \bigcup_{i=1,\dots,n} I_i$, e.g. , $f_{m_1}(I_1,\dots,I_4) = [1,10]$ that represents the union of the intervals of m_1.

The *trade-off operators* lie between conjunctive and disjunctive behaviors, and are typically used when sources are partly conflicting. They try to achieve a good balance between informativeness and reliability [2]. The fusion based on MCS is an example of trade-off operators.

Maximal consistent subset fusion method. When no information is available about sources, like conflict between sources, or reliability of sources, a reasonable fusion method should take into account the information provided by all sources. At the same time, it should try to keep a maximum of informativeness. The notion of MCS is a natural way to achieve these two goals.

The idea of MCS goes back to Rescher and Manor [10]. This notion is currently used in the fusion of logical formulas [5] but also of numerical data [8,9]. Given a set of n intervals $\mathbb{I} = \{I_1, I_2, \dots, I_n\}$, a subset $K \subseteq \mathbb{I}$ is *consistent* if $\bigcap_{i=1}^{|K|} K_i \neq \emptyset$ with $K_i \in K$ and *maximal* if it does not exist a proper super-set $K' \supseteq K$ that is also consistent. In Table 1, the set $K_1 = \{I_1, I_2\}$ is a MCS of the set $\mathbb{I}_{m_1}$, since $I_1 \cap I_2 \neq \emptyset$ and is maximal w.r.t. intersection property.

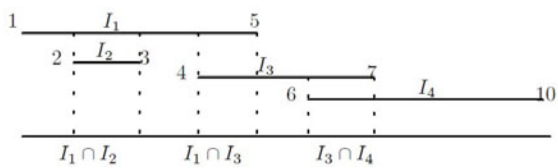

Fig. 2. MCS computed from Table 1 for the variable m_1

The fusion operator of n sources based on MCS consists in applying a disjunctive operator on their MCS. For example, the MCS fusion result for m_1 in Table 1 is $f_{m_1}(I_1,\dots,I_4) = [2,3] \cup [4,5] \cup [6,7]$, as illustrated in Figure 2. The MCS notion appears as a natural way to conciliate the two objectives of gaining information and of remaining in agreement with all sources in information fusion problem. Generally, finding MCS is a problem having exponential complexity [11]. Dubois et al. [8] introduce a linear algorithm to compute the MCS of n intervals.

Properties of fusion operators. Generally, all fusion operators are commutative and idempotent. The conjunctive and disjunctive operators are associative but not the trade-off fusion operators (more details in [9]). If the final result of the fusion is not convex, it is always possible to take its convex hull (loosing some information in the process but gaining computational tractability). Conjunctive fusion result is convex but this is not the case for the others operators in general.

In conclusion, for merging numerical information, a common fusion operator has to be used. This is specially important in case of heterogeneous sources. Fusion operators are often based on assumptions or on meta-knowledge available about the sources (reliability, conflict) and the domain. Sometimes, it happens that the fusion result is not directly useful for decision. For example, in [12] the fused information must be convex, and the convexification of MCS leads to an imprecise result. Here, we propose to identify and characterize interesting subsets of sources, providing more useful fused information. Accordingly, we show how a fusion operator can be embedded in the framework of Formal Concept Analysis (FCA) to build a concept lattice yielding a structured view of partial results labelled by subsets of sources, instead of providing a unique fusion result.

3 Formal Concept Analysis

3.1 Basics

Formal concept analysis (FCA) [13] starts with a formal context (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[1]. The statement $(g, m) \in I$ is interpreted as "the object g has attribute m". An example of formal context is given by Table 2 where a table entry contains a cross ($\times$) iff the object in row has the attribute in column, e.g. g_1 has the attribute m_1, i.e. $(g_1, m_1) \in I$. The two operators $(\cdot)'$ define a Galois connection between the powersets $(2^G, \subseteq)$ and $(2^M, \subseteq)$, with $A \subseteq G$ and $B \subseteq M$:

$$A' = \{m \in M \mid \forall g \in A : gIm\} \qquad B' = \{g \in G \mid \forall m \in B : gIm\}$$

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. $(\{g_1, g_2, g_3\}, \{m_1\})$. In (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 $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$ ($\Leftrightarrow B_2 \subseteq B_1$), e.g. the concept $(\{g_3\}, \{m_1, m_2\})$ is a sub-concept of $(\{g_1, g_2, g_3\}, \{m_1\})$. 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[2] associated with the context in 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 is 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

[1] In this paper, we similarly use the terms object and information source on one hand, and variable and attribute on the other hand.

[2] The lattice diagram is designed with ConExp, `http://conexp.sourceforge.net/`.

intents of its super-concepts. The top concept ($\top$) is the highest and the bottom concept ($\perp$) is the lowest in the lattice.

The concept lattice provides a classification of objects in a domain. It entails both notions of maximality and generalization/specialization: a concept corresponds to a maximal set of objects (extent) sharing a common maximal set of attributes (intent) ; the generalization/specialization is given by the partial ordering of concepts.

However, real-world data like in biology, agronomy, etc., are not binary, but rather consist in complex data composed of numbers, graphs, etc. The data are classically processed with FCA after a data transformation, called *conceptual scaling*, e.g. discretization. Transformations generally imply an important loss of information and arbitrary choices, which must be avoided in the context of information fusion. For example, an object has the attribute m_1 (resp. m_2) in the binary Table 2 iff its values for this attribute are less than 7 (resp. greater than 5) in the numerical Table 1. With other choices, we may obtain another table, and hence another concept lattice with a different interpretation. Therefore, handling numerical data for information fusion purposes with FCA is not straightforward.

3.2 Pattern Structures for Complex Data

Instead of transforming data, one may directly work on the original data. For that purpose, a *pattern structure* is defined as a generalization of a formal context to complex data [14]. It still maps objects to their descriptions, the latter being partially ordered. When working with classical FCA, the object descriptions are sets of attributes, and are partially ordered by set inclusion, w.r.t. set intersection: let $P, Q \subseteq M$ two attributes sets, then $P \subseteq Q \Leftrightarrow P \cap Q = P$, and $(M, \subseteq)$, also written $(M, \cap)$, is a partially ordered set of object descriptions. The set intersection $\cap$ behaves as a meet operator, denoted by $\sqcap$, in a semi-lattice: it is *idempotent*, *commutative*, and *associative*. Therefore, a pattern structure naturally entails a Galois connection between the powerset of objects $(2^G, \subseteq)$ and a meet-semi-lattice of descriptions denoted by $(D, \sqcap)$.

Formally, let G be a set of objects, let $(D, \sqcap)$ be a meet-semi-lattice of potential object descriptions and let $\delta : G \longrightarrow D$ be a mapping. 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 $c, d \in D$ one has $c \sqsubseteq d \Longleftrightarrow c \sqcap d = c$. A pattern structure $(G, (D, \sqcap), \delta)$ gives rise to the following derivation operators $(\cdot)^\square$, given $A \subseteq G$ and $d \in (D, \sqcap)$:

$$A^\square = \bigsqcap_{g \in A} \delta(g) \qquad\qquad d^\square = \{g \in G | d \sqsubseteq \delta(g)\}$$

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), $A \subseteq G$, $d \in (D, \sqcap)$, such that $A^\square = d$ and $A = d^\square$. For a pattern concept (A, d), d is called a *pattern intent* and is a common description of all objects in A, called *pattern extent*. When 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 complete lattice called a *(pattern) concept lattice*.

Pattern structures allow to consider complex data in full compliance with the FCA formalism. It requires to define a meet operator on object descriptions, inducing their partial order. In fact, as for scaling in classical FCA, the choice of an operator depends on expert knowledge, and to which extent will the resulting concept lattice be used. Several attempts were done to define such operators, on sets of graphs [14], numerical data [15], logical formulas [16], etc. In the following, we discuss how a fusion operator can be seen as a meet operator.

4 Organizing Information Fusion Results with FCA

We show here that FCA provides a suitable framework for organizing sources and their information fusion results, allowing more flexibility for the users of fusion results.

Definition (Information fusion space). *An information fusion space D_m is composed of the information available for a variable m and all their possible fusion results, w.r.t a fusion operator f_m.*

For example, with the variable m_1 in Table 1 and f_m as the interval intersection, $D_m = \{[1,5], [4,7], [6,10], [2,3], [4,5], [6,7], \emptyset\}$.

4.1 Formalizing a Fusion Operator as a Meet Operator

Let us consider a single variable $m \in M$, its fusion space D_m corresponding to a chosen fusion operator f_m. When f_m is idempotent, commutative and associative, (D_m, f_m) is a meet-semi-lattice, since f_m behaves as a meet operator. This is the case for any conjunctive or disjunctive fusion operator, and we have $c \sqcap d = f_m(c,d), \forall c, d \in D_m$, meaning that the meet of two elements of D_m corresponds to their fusion.

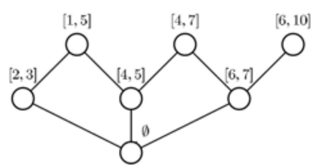

Fig. 3. A meet-semi-lattice of intervals

For example, let us consider the numerical variable m_1 in Table 1, and the conjunctive fusion operator f_{m_1} that corresponds to the interval intersection $\cap$. Figure 3 shows the meet-semi-lattice (D_{m_1}, f_{m_1}). The interval labelling a node is the meet of all intervals labelling its ascending nodes, i.e. the resulting information fusion w.r.t f_{m_1} of the sources given the intervals labelling its ascending nodes. In the example, $f_{m_1}([4,7],[6,10]) = [6,7]$ is the fusion of objects g_3 and g_4 for the variable m_1, and $f_{m_1}([2,3],[1,5]) = [2,3]$ for objects g_1 and g_2. Therefore, we have partially ordered the fusion space D_{m_1} with $c \cap d = c \Leftrightarrow c \subseteq d, \forall c, d \in D_{m_1}$. This order is a particular instance of the pattern subsumption relation defined in pattern structures. It means, in this example, that an interval is subsumed by any larger one, e.g. $[2,3] \sqsubseteq [1,5]$ since $[2,3] \subseteq [1,5]$. For example, we have $[2,3] \sqcap [1,5] = [2,3] \Leftrightarrow [2,3] \sqsubseteq [1,5]$ in terms of semi-lattice, corresponding to $[2,3] \cap [1,5] = [2,3] \Leftrightarrow [2,3] \subseteq [1,5]$ in interval inclusion terms. Note that a disjunctive fusion operator is handled similarly.

4.2 Building a Concept Lattice from Information Sources

Given G a set of sources, $m \in M$ a single variable, (D_m, f_m) the meet-semi-lattice of fusion results, and δ a mapping that gives to any object its information for the variable m, then $(G, (D_m, f_m), \delta)$ is a pattern structure. On the example, we have $(G, (D_{m_1}, f_{m_1}), \delta)$. (D_{m_1}, f_{m_1}) is described in the previous subsection. Descriptions of sources g_1 and g_2 are respectively $\delta(g_1) = [1,5]$ and $\delta(g_2) = [2,3]$. Then, the general Galois connection can be used to compute and order concepts:

$$\begin{aligned} \{g_1, g_2\}^\square &= [1,5] \sqcap [2,3] \\ &= f_{m_1}([1,5],[2,3]) \\ &= [2,3] \end{aligned} \qquad \begin{aligned} [2,3]^\square &= \{g \in G \mid [2,3] \sqsubseteq \delta(g)\} \\ &= \{g \in G \mid [2,3] \subseteq \delta(g)\} \\ &= \{g_1, g_2\}. \end{aligned}$$

Since $\{g_1, g_2\} = [2,3]$ and $[2,3]^\square = \{g_1, g_2\}$, the pair $(\{g_1, g_2\}, [2,3])$ is a concept. Efficient FCA algorithms can extract the set of all formal concepts and order them within a concept lattice [17]. They can be easily adapted to compute in pattern structures [14, 15]. The lattice of our example is given in Figure 4.

4.3 Concept Lattice Interpretation

A concept (A, d) of $(G, (D_{m_1}, f_{m_1}), \delta)$, is interesting from many points of view, as illustrated with the concept $(\{g_1, g_2\}, [2,3])$.

- Its intent d provides the fusion resulting from objects in A, e.g. $[2,3]$ is the conjunctive fusion f_{m_1} of the information from sources g_1 and g_2.
- No other object can be added to A without changing d, e.g. $\{g_1, g_2\}$ is the maximal set of sources whose conjunctive information fusion is $[2,3]$.
- The extent A keeps the track of the origin of the information, e.g. it is known that the new information $[2,3]$ comes from the information of g_1 and g_2.

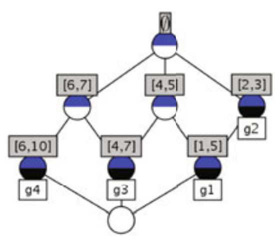

Fig. 4. A concept lattice raised from Table 1 for the variable m_1

The resulting concept lattice provides a suitable classification of information sources and their information fusion results. In Figure 4, a concept extent is read with reduced labelling. However, for sake of readability, intents are given for each concept (not reduced). For example, the node labelled with $[6,7]$ represents the concept $(\{g_3, g_4\}, [6,7])$. Due to concept ordering, a concept provides the fusion result of a subset of the extent of its super-concepts (generalization/specialization). Then, the navigation in the lattice gives interesting insights into the fusion results. This allows more flexibility for decision making. For example, in related works, only the fusion of information of all objects is considered which corresponds to the most general concept ($\top$) in the lattice. This result does not always allow to make a decision, e.g. an empty intersection in our example. Then it is interesting to observe subsets of objects, by navigating in the lattice.

4.4 Lattice Based on Maximal Consistent Subsets

The fusion operator f_m based on the notion of MCS is idempotent and commutative, but not associative. For example in Table 1, $f_{m_1}(f_{m_1}([1,5],[2,3]),[4,7]) = [2,3] \cup [4,7]$ and $f_{m_1}(f_{m_1}([1,5],[4,7]),[2,3]) = [2,3] \cup [4,5]$. Then, the fusion operator cannot be directly used as a meet operator to build a concept lattice.

However, since this operator returns the union of all MCS, we can firstly compute all MCS for a given variable, denoted by the set K and then use the disjunctive operator on the MCS as a meet operator to define a meet-semi-lattice $(K, \cup)$. Formally, we consider $(\mathcal{O}, (K, \cup), \delta)$ as a pattern structure where $\mathcal{O}$ is a multi-set of sources, each element is set of sources of one MCS $k \in K$, i.e. $\delta(o) \in K, \forall o \in \mathcal{O}$. For example, the MCS of intervals for m_1 are $[2,3]$, $[4,5]$ and $[6,7]$ given respectively by $\{g_1, g_2\}$, $\{g_1, g_3\}$ and $\{g_3, g_4\}$. Then, $\mathcal{O}$ represents the multi-set $\{\{g_1, g_2\}, \{g_1, g_3\}, \{g_3, g_4\}\}$ with $\delta(\{g_1, g_2\}) = [2,3]$ (meaning that the interval of values $[2,3]$ is related to the sources g_1 and g_2), $\delta(\{g_1, g_3\}) = [4,5]$ and $\delta(\{g_3, g_4\}) = [6,7]$. Then, we use an interval union as a meet operator. The resulting concept lattice is given in Figure 5. A concept extent is read with reduced labelling. A concept intent is given here for each concept. For example, in Figure 5, the right concept in the second line is $(\{\{g_1, g_2\}, \{g_1, g_3\}\}, [2,3] \cup [4,5])$ giving the values of m_1 w.r.t. the sources $\{g_1, g_2\}$ and $\{g_1, g_3\}$. Moreover, these values represent the MCS fusion result of the subset $\{g_1, g_2, g_3\}$. The concept $\top$ corresponds to the union of all MCS that is the MCS fusion result of all sources.

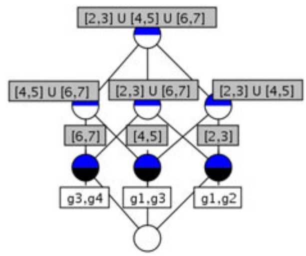

Fig. 5. Concept lattice with MCS

The method used here to obtain the lattice based on MCS does not consider all subsets of objects with their MCS fusion results. This is due to the non-associativity of the MCS fusion operator. Thus, the concept lattice does not contain all subsets of G with their MCS fusion results since the interval union is used on the MCS of data and not directly on the data given by sources. Nevertheless, the concept lattice helps us to keep the origin of the information and gives more flexibility for the users in the choice of a maximal consistent subset of sources in many application fields.

4.5 Embedding Several Variables in the Concept Lattice

Sources can provide values for different variables. For example, Table 1 involves objects described by vectors of intervals, where each dimension, i.e. column, corresponds to a unique variable, e.g. the description of the object g_1 is denoted by $\delta(g_1) = \langle [1,5], [1,9] \rangle$. It can be interesting to compute the fusion information for all variables simultaneously.

To formalize a pattern structure in this case, one defines a meet operator, i.e. fusion operator in our settings, for each dimension, or variable. Assuming that there is a canonical order on vector dimensions, the meet of two vectors is defined as the meet on each dimension. This induces a partial order of object descriptions [15]. Thus, we consider the pattern structure $(G, (D, \sqcap), \delta)$, where G is a set of sources, $(D, \sqcap)$ is a meet-semi-lattice of vectors, and each vector dimension is provided with the fusion operator f_m corresponding to the variable m.

Going back to Table 1, descriptions of objects g_1 and g_2 are respectively the vectors $\langle[1, 5], [1, 9]\rangle$ and $\langle[2, 3], [1, 3]\rangle$. When the fusion operator for both dimension is the interval intersection, the meet of these two vectors is $\langle[1, 5], [1, 9]\rangle \sqcap \langle[2, 3], [1, 3]\rangle = \langle[2, 3], [1, 3]\rangle$. The subsumption relation for vectors is defined similarly: $\langle[2, 3], [1, 3]\rangle \sqsubseteq \langle[1, 5], [1, 9]\rangle$ as $[2, 3] \subseteq [1, 5]$ and $[1, 3] \subseteq [1, 9]$. Then, the general Galois connection can be used to compute and order concepts:

$$\{g_1, g_2\}^\square = \langle[1, 5], [1, 9]\rangle \sqcap \langle[2, 3], [1, 3]\rangle = \langle[2, 3], [1, 3]\rangle \qquad \langle[2, 3], [1, 3]\rangle^\square = \{g \in G \mid \langle[2, 3], [1, 3]\rangle \sqsubseteq \delta(g)\} = \{g_1, g_2\}$$

In this way, a concept represents a set of sources and their fusion w.r.t. all variables, such as no other source can be added without changing the fusion result for any variable. The variables can be either symbolic or numerical since a fusion operator is chosen for each variable.

When the fusion operator is based on MCS, we follow the pre-processing introduced above for each variable (see Section 4.4). Then, we consider the set of all MCS for all variables. Thus, we consider the pattern structure $(\mathcal{O}, (K, \sqcap), \delta)$, where $\mathcal{O}$ is the set of subsets of sources providing the MCS for all variables, $(K, \sqcap)$ is a meet-semi-lattice of vectors. Each subset in $\mathcal{O}$ is described for each dimension by a maximal interval of values if the subset represents a MCS for the corresponding dimension, otherwise the dimension description is empty. In the example, recalling that an object denotes a set of sources giving a MCS, the description of the object $\{g_1, g_2\}$ is $\delta(\{g_1, g_2\}) = \langle[2, 3], [1, 3]\rangle$ where $[2, 3]$ and $[1, 3]$ are respectively a MCS for m_1 and m_2. By contrast, the description of the object $\{g_3, g_4\}$ is $\delta(\{g_3, g_4\}) = \langle[6, 7], \emptyset\rangle$ since the subset $\{g_3, g_4\}$ does not represent a MCS for the variable m_2.

This framework on fusion operators has been used on real-world data as explained in the next section.

5 A Real-World Application in Agronomy

Data and problem settings. Agronomists compute indicators for evaluating the impact of agricultural practices on the environment. Questions such as the following are of importance: what are the consequences of the application of a pesticide given its characteristic, the period of application, and the characteristics of the field? The risk level for a pesticide to reach groundwater is computed by the indicator I_{gro} in [18]. Agronomists try to make a diagnosis w.r.t. the value of I_{gro}. A value below 7 indicates that the farmer has to change its practices (pesticide, soil, date, etc.). By contrast, a value above 7 indicates that the practices of the farmer are environmental friendly [19]. Pesticide characteristics depend on the chemical characteristics of the product while pesticide period application and field characteristics depend on domain knowledge [19]. This knowledge lies in information sources among which books, databases, and expert knowledge in agronomy. Then values for some characteristics vary w.r.t. sources. Here, we are interested in the use of pesticide *sulcotrione* and its influence on the groundwater. Sulcotrione is a herbicide marketed since 1993. It is used to control a wide range of grasses weeds in maize crops. Sulcotrione is generally weakly absorbed by soils [20]. Three characteristics of *Sulcotrione* are needed to compute the indicator I_{gro}, namely $DT50$, *koc*, and ADI (more details on these characteristics can be found in [18], and are not crucial for the understanding of this paper). Table 3 (simplified data) gives the values of the characteristics $DT50$ and *koc* according to 9 different information sources. The symbol "?" represents the case when the information source does not give data for the characteristic. The value of ADI for the *sulcotrione* is 0.00005. Agronomists look to find a suitable value for each characteristic to be considered for computing the I_{gro} indicator, hence facing an information fusion problem.

Table 3. Characteristics of *Sulcotrione*

	DT50 day	koc L/kg
BUS	[2,74]	?
PM11	[15,72]	?
PM12	?	[44,940]
PM13	?	[44,940]
INRA	?	[1.08,8.98]
Com98	[2,6]	[17,160]
AGXf	[2,6]	[1.08,160]
AGXl	[15,74]	[1.08,160]

Lattice construction and interpretation. To combine the different pieces of information, a common fusion operator has to be defined. In this application, (1) the sources are heterogeneous (2) no *a priori* knowledge about sources and characteristics is available. Therefore, an appropriate fusion operator is the MCS fusion operator. The MCS for the variable $DT50$ are K_1 and K_2, resp. K_3 and K_4 for *koc* (see Table 4). Table 5 results from the pre-processing of Table 3, detailed in Section 4.4. The resulting concept lattice is given in Figure 6 with 16 concepts. A concept extent is read with reduced labelling. A concept intent is not given in vectorial form for sake of readability: it is read from the intents of sub-concepts, for example, the intent of the concept C_1 is $\{(DT50, [15, 72]), (koc, [44, 160])\}$. But, if two sub-concepts intents give different values for a same attribute, then the union of values is considered. For example, the intent of the concept C_2 is $\{(DT50, [2, 6] \cup [15, 72]), (koc, [44, 160])\}$ and its sub-concepts intents are $\{(DT50, [2, 6])\}$, $\{(DT50, [15, 72])\}$ and $\{(koc, [44, 160])\}$. Moreover, each concept intent in the lattice represents the MCS fusion result of the subset of sources in the extent. The highest concept in the lattice corresponds to the MCS fusion result of all sources for all characteristics. For example, the "most right-down" concept is $(\{K_1\}, \{(DT50, [2, 6])\})$ where $[2, 6]$ is the MCS fusion result of the subset $K_1 = \{BUS, Com98, AGXf\}$ and its "most right" super-concept is $(\{K_1, K_2\}, \{(DT50, [2, 6] \cup [15, 72])\}$ where $[2, 6] \cup [15, 72]$ is the fusion result of the set $K_1 \cup K_2 = \{BUS, PM11, AGXl, Com98, AGXf\}$.

Results and discussion. The computing of a lower and higher bound for the indicator and the consequences of the results on agronomic practices and pollution are detailed and discussed in [12], but will not be detailed here as this is not necessary. It is required to consider the convex hull of the fusion result for computing the indicator. The concept lattice allows the users of I_{gro} and experts to give several diagnosis for the farmer. For example, let us consider the concept

Table 4. Label of all MCS

K_1	$\{BUS, Com98, AGXf\}$
K_2	$\{BUS, PM11, AGXl\}$
K_3	$\{INRA, Com98, AGXf, AGXl\}$
K_4	$\{PM12, PM13, Com98, AGXf, AGXl\}$

Table 5. Table 3 pre-processed

	DT50 (days)	koc (L/kg)
K_1	[2,6]	$\emptyset$
K_2	[15,72]	$\emptyset$
K_3	$\emptyset$	[1.08,8.98]
K_4	$\emptyset$	[44,160]

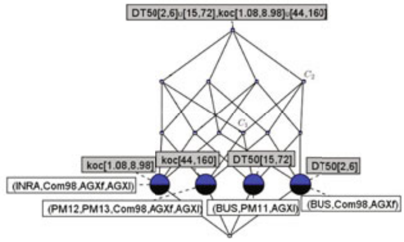

Fig. 6. Concept lattice built from Table 5

$\top$ that represents the fusion result of all sources for all characteristics. Then, $DT50$ and koc lie respectively in $[2, 72]$ and $[1.08, 160]$. With these values, the computed value for I_{gro} is $[4, 10]$. This interval is not useful since all values in $[4, 10]$ are neither smaller than 7 nor greater than 7 and the expert cannot make a decision on the practices of the farmer.

Now the indicator I_{gro} can be also computed choosing either intervals of values in higher or lower level concepts. For instance, if we consider the values of $DT50$ in $[2, 6]$, koc in $[44, 160]$ then we obtain the interval $[9.97, 10]$ for I_{gro} and the practices of the farmer are environmental friendly since the I_{gro} value is greater than 7. However, if $DT50 = [15, 72]$ and $koc = [1.08, 8.98]$, the resulting interval for I_{gro} is $[4.32, 4.32]$ indicating that the farmer must change its practices since values of I_{gro} are smaller than 7. Anyhow, we obtain, with these concepts, precise results of I_{gro}, which is not the case with the fusion global result when using the most general concept. The concept lattice allows to identify what maximal subsets of sources support the most precise results. A further step is to consider these precise results in a decision process.

6 Conclusion

In this paper, we claim that Formal Concept Analysis has the capability of supporting a decision making process in the presence of information fusion problems, even when information are complex, e.g. numbers, thanks to the formalism of pattern structures. A real-world experiment in agronomy showed that when a fusion result does not allow to make a decision, the concept lattice helps the expert by considering an ordered hierarchy of concepts, given the fusion from different maximal sets of sources. Some fusion operators can directly be used to build a concept lattice, e.g. conjunctive and disjunctive operators. To deal with the operator based on maximal coherent subsets (MCS), we proposed to transform the data since MCS is not an associative operator, and the resulting concept lattice entails fusion results of interest.

We have considered the case when information are represented by fuzzy intervals and possibility distributions in [21]. As perspective, it is interesting to study how other fusion operators can be embedded in a concept lattice, as well as meta-information on sources (when available).

References

1. Bloch, I., Hunter, A., Ayoun, A., Benferhat, S., Besnard, P., Cholvy, L., Cooke, R., Dubois, D., Fargier, H.: Fusion: general concepts and characteristics. International Journal of Intelligent Systems 16, 1107–1134 (2001)
2. Dubois, D., Prade, H.: Possibility theory and data fusion in poorly informed environments. Control Eng. Practice 2, 811–823 (1994)
3. Dubois, D., Lang, J., Prade, H.: Dealing with multi-source information in possibilistic logic. In: European Conference on Artificial Intelligence, pp. 38–42 (1992)
4. Dubois, D., Prade, H.: Possibility theory in information fusion. In: Data fusion and Perception, pp. 53–76 (2001)
5. Benferhat, S., Dubois, D., Prade, H.: Reasoning in inconsistent stratified knowledge bases. In: Int. Symp. on Multiple-Valued Logic, pp. 184–189 (1996)
6. Chaudron, L., Maille, N.: Le modèle des cubes: représentation algébrique des conjonctions de propriétés. In: Reconnaissance des Formes et Intelligence Artificielle, RFIA (2000) (in French)
7. Phan-Luong, V.: A framework for integrating information sources under lattice structure. Information Fusion 9(2), 278–292 (2008)
8. Dubois, D., Fargier, H., Prade, H.: Multiple source information fusion: a practical inconsistency tolerant approach. In: IPMU, pp. 1047–1054 (2000)
9. Destercke, S., Dubois, D., Chojnacki, E.: Possibilistic information fusion using maximal coherent subsets. IEEE Transactions on Fuzzy Systems 17, 79–92 (2009)
10. Rescher, N., Manor, R.: On inference from inconsistent premisses. Theory and Decision 1, 179–217 (1970)
11. Malouf, R.: Maximal consistent subsets. Comput. Linguist. 33(2), 153–160 (2007)
12. Assaghir, Z., Girardin, P., Napoli, A.: Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment. In: IFSA/EUSFLAT Conf., pp. 707–712 (2009)
13. Ganter, B., Wille, R.: Formal Concept Analysis. Springer, Heidelberg (1999)
14. Ganter, B., Kuznetsov, S.O.: Pattern Structures and Their Projections. In: Delugach, H.S., Stumme, G. (eds.) ICCS 2001. LNCS (LNAI), vol. 2120, pp. 129–142. Springer, Heidelberg (2001)
15. Kaytoue, M., Duplessis, S., Kuznetsov, S.O., Napoli, A.: Two FCA-Based Methods for Mining Gene Expression Data. In: Formal Concept Analysis. LNCS, vol. 5548. Springer, Heidelberg (2009)
16. Chaudron, L., Maille, N.: Generalized formal concept analysis. In: Ganter, B., Mineau, G.W. (eds.) ICCS 2000. LNCS, vol. 1867, pp. 357–370. Springer, Heidelberg (2000)
17. Kuznetsov, S.O., Obiedkov, S.A.: Comparing Performance of Algorithms for Generating Concept Lattices. J. Exp. Theor. Artif. Intell. 14, 189–216 (2002)
18. Van der Werf, H., Zimmer, C.: An indicator of pesticide environmental impact based on a fuzzy expert system. Chemosphere 36(10), 2225–2249 (1998)
19. Bockstaller, C., Girardin, P., Van Der Werf, H.: Use of agro-ecological indicators for the evaluation of farming systems. European Journal of Agronomy 7(1-3) (1997)
20. Baer, U., Calvet, R.: Fate of Soil Applied Herbicides: Experimental Data and Prediction of Dissipation Kinetics. J. Environ. Qual. 28(6), 1765–1777 (1999)
21. Assaghir, Z., Kaytoue, M., Prade, H.: A possibility theory-oriented discussion of conceptual pattern structures. In: Int. Conf. on Scalable Uncertainty Management (SUM). LNCS. Springer, Heidelberg (2010)