archives-ouvertes.fr

Numerical information fusion: Lattice of answers with supporting arguments

Zainab Assaghir, Amedeo Napoli, Mehdi Kaytoue, Didier Dubois, Henri Prade

Numerical information fusion: Lattice of answers with supporting arguments

Zainab Assaghir, Amedeo Napoli
LORIA
Campus Scientifique, B.P. 70239 – 54506
Vandoeuvre-les-Nancy, France
{assaghiz,napoli}@loria.fr

Mehdi Kaytoue
Universidade Federal de Minas Gerais
Av. Antonio Carlos, 6627
Belo Horizonte, Brazil
{kaytoue}@dcc.ufmg.br

Didier Dubois, Henri Prade
IRIT
118 Route de Narbonne – 31062
Toulouse, France
{dubois,prade}@irit.fr

Abstract—The problem addressed in this paper is the merging of numerical information provided by several sources. Merging conflicting pieces of information into an interpretable and useful format is a tricky task even when an information fusion method is chosen. The use of formal concept analysis and pattern structures enables us to associate subsets of sources to combination results obtainable from consistent subsets of pieces of information. This provides a lattice of arguments where the reliability of sources can be taken into account. Instead of providing a unique fusion result, the method yields a structured view of partial results labelled by subsets of sources and allows us to argue about the most appropriate evaluation. The approach is illustrated with an experiment on a real-world application to decision aid in agricultural practices.

Keywords- numerical information fusion, reliability, formal concept analysis, pattern structure, lattice of arguments.

I. Introduction

In several application domains, information is available from several sources rather than a single one. Extracting consistent and useful pieces of information from such distributed sources is the major task of information fusion [1]. Moreover information delivered by sources may be uncertain or imprecise, and then these aspects have to be taken into account in information fusion. There are several information fusion operators for combining pieces of information that may be often conflicting [2]. These fusion operators are applied on the set of all sources and provide a result.

In this work, we use Formal Concept Analysis (FCA) to draw a map of all subsets of sources and the information they provide [3]. The main capability of FCA is to produce pairs of subsets of totally related entities called formal concepts. Hence, concepts encode maximal sets of sources associated with the result of the fusion process. Concepts are ordered and form a structure called concept lattice. This lattice is meaningful for organizing information fusion results from different subsets of sources and allows more flexibility for the user. Moreover, the lattice keeps track of the origin of the information such as presented in [4] for the fusion of symbolic information. Besides, the paper investigates a method for coping with inconsistency, by defining a suitable notion of *lattice of arguments*, when merging inconsistent information.

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

The paper is organized as follows. Section II introduces the preliminaries on FCA and its extension for handling numerical data. Then, Section III shows how FCA is well suited for organizing different information fusion results. Section IV introduces the notion of lattice of arguments. Then, Section V presents how the reliability of sources can be taken into account. Section VI describes a real-world experiment: a concept lattice embedding fusion results is interpreted for making decisions about agricultural practices.

II. Formal Concept Analysis

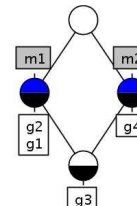|	m_1	m_2
g_1	×	
g_2	×	
g_3	×	×
g_4		×

Table I
A FORMAL CONTEXT

Figure 1. Concept lattice induced by Table I

A. Basics

Formal Concept Analysis (FCA) is a method mainly used for the analysis of data. Implicit relationships between objects can be derived and described through a set of attributes. The FCA formalism is based on lattice structures [3]. FCA computes a *concept lattice* which provides a classification of objects in a domain. FCA starts from a binary table called *formal context* (G, M, R) that represents a relation R between a set of objects G and a set of attributes M. The statement $(g, m) \in R$ is interpreted as "object g has attribute m". An example of formal context is given on Table I where a table entry contains a cross (×) iff the object in the corresponding row possesses the attribute in the corresponding column, e.g. g_1 has attribute m_1, i.e.

$(g_1, m_1) \in R$. The two operators $(\cdot)'$ define a Galois connection between the powersets $(2^G, \subseteq)$ and $(2^M, \subseteq)$, with $A \subseteq G$ and $B \subseteq M$:

$$A' = \{m \in M \mid \forall g \in A : gRm\} \qquad for\ A \subseteq G,$$
$$B' = \{g \in G \mid \forall m \in B : gRm\} \qquad for\ B \subseteq M$$

A pair (A, B), such that $A' = B$ and $B' = A$, where $A \subseteq G$, $B \subseteq M$, is called a *(formal) concept*, e.g. $(\{g_1, g_2, g_3\}, \{m_1\})$. A concept includes 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 set A of objects is called the *extent* and the set B of properties is the *intent* of the concept (A, B).

Concepts are then partially ordered to form a lattice. The generalization/specialization is given by the partial ordering of concepts. 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, R). Figure 1 shows the concept lattice[1] associated with the context in Table I. 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 intents of its super-concepts. The top concept $(\top)$ of the lattice contains all objects and the bottom concept $(\bot)$ in the lattice contains all properties. Thus, FCA allows us to associate a maximal subset of objects with a maximal subset of attributes.

Real-world data in biology, agronomy, etc., usually consist in complex data involving numbers, intervals, graphs, etc. (e.g. Table II). FCA cannot be directly applied to such data. They can be processed with FCA after a transformation, called *conceptual scaling*, e.g. discretization. Transformations generally imply an important loss of information (e.g. links between objects) and arbitrary choices. However, there are means to handle numerical values and intervals in an elegant and efficient way in FCA. For that purpose, Ganter and Kuznetsov [5] define *pattern structures* that directly handle complex data such as given in Table II.

	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 II
INFORMATION ITEM INTERVALS

[1]Lattice diagram designed with ConExp, http://conexp.sourceforge.net/.

B. Pattern structures for complex data

Pattern structures are introduced in [5] as an extension of FCA to handle complex data. When working with classical FCA, the object descriptions are sets of attributes, and are partially ordered by set inclusion. But inclusion is the canonical ordering on sets, induced by set-intersection: let $P, Q \subseteq M$ be two attribute sets, then $P \subseteq Q \Leftrightarrow P \cap Q = P$. So the partially ordered set of attributes $(M, \subseteq)$ can as well be written $(M, \cap)$. Based on this remark, a meet operator denoted by $\sqcap$, idempotent, commutative and associative, can be applied on the set of object descriptions and the potential descriptions are ordered in a meet-semi-lattice w.r.t. the induced ordering $\sqcap$. Therefore, a pattern structure 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 a subsumption relation $\sqsubseteq$ such that 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 $d \in (D, \sqcap)$:

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

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 the 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 us to consider complex data in full compliance with FCA formalism. It requires the definition of a meet operator on object descriptions, inducing their partial order. Actually, as for scaling in classical FCA, the choice of an operator depends on expert knowledge, and the context in which the resulting concept lattice will be used.

III. INFORMATION FUSION USING FCA

In information fusion, the question is to synthesize several pieces of information, providing by different sources, into an interpretable and useful information. Pieces of numerical information supplied by sources are often imprecise. The simplest representation of this imprecision consists of intervals. For example, Table II represents an example of information fusion problem : how to synthesize a useful value for m_1 (resp. for m_2), based on values provided by information sources g_i, $i = 1, \ldots, 4$, to be used in an evaluation process. According to [1], [2], there are three kinds of behaviors for a fusion operator denoted f: conjunctive, disjunctive and mild operators.

- The basic *conjunctive operator* is a set intersection. It is used when all the sources are reliable, and usually results in a more precise information item. If there is some conflict in the information, then the result of the conjunction can be empty. For example, considering Table II, the conjunctive fusion result is an emptyset for the variable m_1.
- The basic *disjunctive operator* is a set union. It is used when at least one of the sources is reliable without knowing which sources are so. The result of a disjunctive operator can be considered as reliable, but is also often imprecise. For example, the disjunctive fusion result for the variable m_1, in Table II, is $[1, 10]$.
- A *mild operator* lies between conjunction and disjunction, and is typically used when sources are partly conflicting. It tries to achieve a good balance between informativeness and reliability. The fusion based on *Maximal Consistent Subset (MCS)* is an example of mild operator, which consists in finding maximal consistent conjunctions of sources, then returning the disjunction of these maximal sets [6], [7]. Going back to Table II, maximal consistent subsets of values for m_1 are $[2, 3]$, $[4, 5]$ and $[6, 7]$ provided respectively by subsets $\{g_1, g_2\}$, $\{g_1, g_3\}$ and $\{g_3, g_4\}$. Then, the MCS fusion result is given by $[2, 3] \cup [4, 5] \cup [6, 7]$.

Conjunctive and disjunctive fusion operators are commutative, associative and idempotent. The mild operators are commutative and idempotent but they are not associative. Nevertheless, fusion operators are generally applied to all sources and the obtained result (called also *global* result) is considered by users for the evaluation process. Sometimes, it happens that this result is not useful, e.g. the intersection of all sources in Table II is an emptyset. Then, we are interested to consider subsets of pieces of information provided by sources. Thus, it may be useful to keep sources distinct and exploit the *partial* fusion results (i.e. results obtained from subsets of sources). Nevertheless, there exist 2^n subsets of sources for a set of n sources. On the other hand, FCA (in particular pattern structure) starts from a similar kind of information as in information fusion problem table. Moreover, it produces concepts corresponding to maximal sets of objects associated with a maximal set of values [8]. Then, FCA (i.e. pattern structures for numerical data) can be used to combine numerical information coming from several sources. Indeed, information sources (resp. variables) in information fusion problems play the role of objects (resp. attributes) in the pattern structure formalism. Moreover a fusion operator is considered as a meet operator in pattern structures and FCA formalism. Table III represents some related basic notions in FCA, pattern structures and information fusion that we will detail later in this paper to show how FCA and pattern structures are well suited for merging information.

FCA	Pattern structures	Information fusion
objects	objects	sources
attributes	descriptions	variables
binary data	complex data	intervals
meet operator	meet operator $\sqcap$	fusion operator f_m
	D description space	D_m fusion space
$(2^M, \cap)$	$(D, \sqcap)$	(D_m, f_m)
$(2^M, \subseteq)$	$(D, \sqsubseteq)$	$(D_m, \sqsubseteq_{f_m})^a$
formal concept	pattern concept	argument
concept lattice	pattern concept lattice	lattice of arguments

$^a \sqsubseteq$ and $\sqsubseteq_{f_m}$ are respectively defined with respect to $\sqcap$ and f_m. For instance $\sqsubseteq_{f_m}$ is the interval inclusion if f_m is a conjunctive operator.

Table III
RELATED NOTIONS IN THE FIELDS OF FCA, PATTERN STRUCTURES AND INFORMATION FUSION

Let us consider a set of sources $G = \{g_1, g_2, \ldots, g_n\}$. Each source g_i provides an information item in the form of an interval $[a_i, b_i]$ for a parameter m. f_m a fusion operator and $f_m(g_i, g_j)$ is the fusion result of information items coming from g_i and g_j.

Definition 1: An information fusion space D_m is composed of the information available for a parameter m and all possible fusion results they provide, w.r.t a fusion operator f_m. Formally, $D_m = \{f_m(g_i \in H), \forall H \subseteq G\}$.
Going back to Table II, with variable m_1 and f_1 being the conjunctive fusion operator (i.e. interval intersection), $D_{m_1} = \{[1, 5], [4, 7], [6, 10], [2, 3], [4, 5], [6, 7], \emptyset\}$.

A. Lattice of conjunctive results

The conjunctive fusion operator $f_m = \cap$, for a parameter m, is commutative, associative and idempotent. Then, it can be considered as a meet operator and $(D_m, \cap)$ is an information meet-semi-lattice. Consequently, $(G, (D_m, \cap), \delta)$ is a pattern structure and we can build the lattice given in Figure 2.

On Figure 2, from the FCA point of view, each node denotes a concept while a line denotes an order relation between two concepts (i.e., interval inclusion). In information fusion problems, each node contains a subset of sources (extent in FCA) and an interval (intent in FCA) corresponding to the intersection (i.e. the meet in FCA) of all intervals given by sources. The top node ($\top$) in the lattice corresponds to all sources and the bottom node ($\bot$) to no source at all.

The lattice gives a meaningful map of the available information and the fusion results. Each node contains a maximal subset of sources and the information fusion result it supplies. Moreover, the lattice provides a structured view of global and partial conjunctive fusion results. Actually, a pair (A, d) of $(G, (D_{m_1}, \cap), \delta)$, is interesting from many points of view, as illustrated with the node labelled $(\{g_1, g_2\}, [2, 3])$.
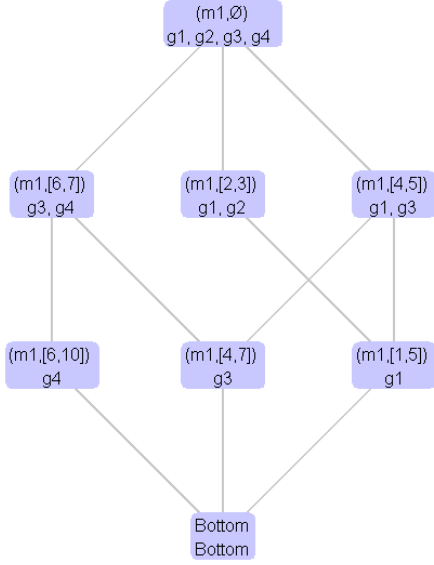
Figure 2. Conjunctive results lattice

- The value d provides information fusion resulting from sources in A, e.g. $[2,3]$ is the conjunctive fusion of information coming from sources g_1 and g_2.
- No other source 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 subset A keeps the track of the origin of information, e.g. it is known that the new information $[2,3]$ comes from g_1 and g_2.
- The top node ($\top$) in the lattice gives the global fusion result (i.e. fusion result of all sources in G).

Moreover, when the global fusion result is conflicting hence not useful (for example, in the conjunctive fusion lattice, the global fusion result is the empty set), the lattice allows users to choose several subsets of sources with their fusion results that are the more useful and the corresponding non-empty intervals. Then, it is interesting to observe subsets of sources given in the lattice and their partial fusion results.

Note that the disjunctive fusion operator allows to build another lattice since it is commutative, associative and idempotent.

Besides, information sources give information about several variables. Then, we must compute one information fusion lattice for each variable. In the pattern structure formalism [9], when the description of objects is representing by a vector of object descriptions for each variable, we consider a meet operator for each attribute. Then, we consider that information fusion sources describe variables using a multi-dimentional vector. Each dimension corresponds to a variable. Consequently, we use a fusion operator for each dimension in the source description. This method

allows us to obtain the fusion result for several variables simultaneously.

B. Lattice of MCS results

The MCS fusion operator cannot be used as a meet operator since it is not associative. For instance, going back to Table II and f_1 represents the MCS fusion operator, $f_1((g_1, g_2), g_3) = [2, 3] \cup [4, 7]$ and $f_1((g_1, g_3), g_2) = [2, 3] \cup [4, 5]$. Then, we cannot directly build the MCS fusion results lattice. However, the MCS fusion operator returns the disjunction of maximal consistent conjunctions. Consequently, we can use the disjunctive operator, on the conjunctions obtained from initial information, as a meet operator to build the lattice of MCS information fusion results. Then, we use a pre-processing method, detailed in [8], that consists in computing maximal consistent subsets of sources (i.e. MCS) with the associated conjunction of intervals; then we apply a disjunctive operator (viewed as a meet operator) on these MCSs and derive another meet-semi-lattice $(D_m, \cup)$.

Then, consider the set K of subsets of sources providing the MCS values and let D_m represent the information fusion space of the parameter m. Consequently, $(K, (D, \cup), \delta)$ is a pattern structure and we can build another lattice.

Going back to Table II, $K = \{K_1, K_2, K_3\}$ where $K_1 = \{g_1, g_3\}$, $K_2 = \{g_3, g_4\}$ and $K_3 = \{g_1, g_2\}$ for m_1. Figure 3 represents the lattice of MCS results for variable m_1. In [10],
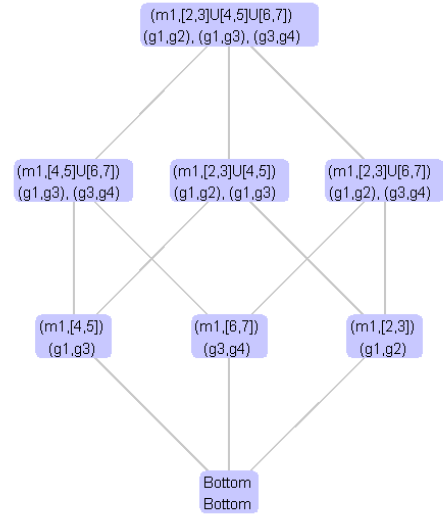

Figure 3. Lattice of MCS fusion results

the fusion result must be convex, and the convex hull of MCS fusion results is imprecise. Therefore, using a subset of sources, i.e. a partial fusion result is more useful and the fusion result is more precise. Hence, the lattice of MCS results organizes the groups of sources with their merged

results. Each node corresponds to a subset of sources with their MCS result. On Figure 3, let us consider the node labelled $(\{(g_1, g_3), (g_3, g_4)\}, (m_1, [4, 5] \cup [6, 7]))$. The value $[4, 5] \cup [6, 7]$, of the variable m_1, is coming from K_1 and K_2. Moreover, these values represent the MCS fusion result of the subset $\{g_1, g_3, g_4\}$. The top node $\top$ corresponds to the union of all MCS of sources that is the MCS fusion result of all sources.

Besides, if the information sources provide globally consistent information, then the conjunction of all items is not empty. In that case, the lattice of MCS results reduces to a unique node (this corresponds to the intersection of all source items since they are consistent). However, if the information sources are pairwise conflicting, then each concept of the conjunctive fusion lattice contains a single source and the pieces of information it provides, the $\top$ and the $\bot$ nodes. More generally, if sources are conflicting, the global conjunctive fusion result is empty. Then, nodes located just below the $\top$ (called *co-atom* concepts) in the conjunctive fusion results lattice, form nodes of the MCS fusion results lattice (called *atom* concepts). Instead of the pre-processing method for the MCS results lattice, we can build the lattice of conjunctive results. Then, it consists in deleting the $\top$ and performing unions to obtain MCS results. Note the obtained structure is based on both intersection and union semi-lattices, but it is not a lattice structure.

IV. Concept lattice as lattice of arguments

The concept lattice gives a classification of sources and information provided by sources. Maximal fusion results are organized in the lattice. When the global fusion result corresponding to the fusion of all sources is empty or imprecise, the lattice allows us to consider partial fusion results that correspond to subsets of sources. Then, the lattice forms a hierarchy of arguments or answers for an evaluation process. Instead of the global fusion each subset of sources with their fusion results in the lattice can be used in the evaluation process. Each node in the lattice provides a decision for the user. Then, the lattice is the basis for answering queries or displaying arguments for several decisions. Each pair containing a subset of sources with their fusion result leads to a possible decision. Hence, the lattice can be called *lattice of answers* or *lattice of arguments*.

Formally, let us consider $S = \{g_1, \ldots, g_k\}, k \leq n$ a subset of sources in G and $f_m(S)$ is the fusion result of information provided by sources in S w.r.t. a fusion operator f_m for the parameter m. Then, the pair $(S, f_m(S))$ is called an argument or an answer. Hence, the concept lattice is said to be a lattice of arguments or a lattice of answers if and only if the subset S is maximal with respect to $f_m(S)$, i.e. for any S_0 such that $f_m(S_0) = f_m(S)$, we have $S_0 \subseteq S$. Thus, the information $f_m(S)$ is associated with a maximal subset of sources that support it.

Instead of providing a unique fusion result, the lattice of arguments gives a structured view of optimal partial fusion results labelled by subsets of sources. This allows us to choose among these subsets of sources and argue about the most appropriate evaluation using the reliability of sources. In the next section, we will detail how the reliability of sources can be taken into account to provide a synthetic conclusion in the evaluation process.

V. Taking into account the source reliability

The reliability of a source represents its capability to provide a correct piece of information for a given problem. There are two approaches to modeling such reliability:

- One may define a reliability complete pre-ordering describing whether one source is more reliable than another or not.
- One is capable to determine a numerical reliability factor or certainty degree for each source, denoted by $\alpha \in [0, 1]$. We can let $\alpha = 1$ to denote that the source is fully reliable and $\alpha = 0$ when the source is totally unreliable (in the sense of useless: the information provided by the corresponding source is not taken into account in the evaluation process).

Reliability can be context-dependent. For example, even if some source is more reliable than others regarding a parameter, these other sources maybe more worthy of trust than the former for another parameter. The question is how to use reliability information in the lattice of arguments and more generally, how to combine information from n sources taking into account the reliability of the sources in order to provide a synthetic conclusion for the user. For a set of n sources, let $\alpha_j > 0$ be the reliability degree of source g_j for a parameter m. When all sources are reliable, the conjunctive fusion operator is used to combine pieces of information and the global evaluation is considered.

In a quantitative uncertainty setting, Shafer's evidence theory [11] provides a way for discounting the information supplied by unreliable sources. The degree of reliability α is interpreted as the probability that the information d provided by a source is relevant and correct. In other words, the degree $1 - \alpha$ is the probability that the information is useless. A pair (d, α) formed by an interval d and a reliability degree α is then interpreted as a consonant belief function defined by a mass function $mass(d) = \alpha$ and $mass(DOM) = 1 - \alpha$ where DOM is the attribute domain. It is equivalent to replace d by a possibility distribution on DOM, such that $\pi(x) = 1$ if $x \in d$ and $1 - \alpha$ otherwise.

As the application of FCA requires idempotent operations we shall use minimum and maximum to combine such possibility distributions. The combination using the *minimum* is in agreement with a logical point of view on the information fusion problem. One can generalize the MCS method to the handling of reliability weights attached to sources for each argument in the lattice of arguments. Suppose one source

is characterized by (d_1, α_1) and another one by (d_2, α_2). Combining the two possibility distributions by the minimum, and assuming $d_1 \cap d_2$ is a non-empty interval, one easily deduces that necessity degrees of $d_1 \cap d_2$ and $d_1 \cup d_2$ are respectively $\min(\alpha_1, \alpha_2)$ and $\max(\alpha_1, \alpha_2)$. However, if there are conflicts between sources, the conjunction of possibility distributions is not sufficient.

One idea is then to use qualitative counterparts [12] of combination rules existing in evidence theory setting. Define a qualitative mass function μ on 2^{DOM} associated with the pair (d, α) as $\mu(d) = \alpha$ and $\mu(DOM) = 1$. The corresponding necessity measure is defined by $N(d_1) = \max_{e \sqsubseteq d_1} \mu(e)$. The conjunctive rule assigns the mass $\min(\alpha_1, \alpha_2)$ to $d_1 \cap d_2 \neq \emptyset$ (and 0 otherwise), α_1 to d_1, α_2 to d_2, and 1 to DOM. When there is no conflict, the result is still a possibility distribution, due to the use of min operation, since assuming $\alpha_1 \leq \alpha_2$, $\min(\alpha_1, \alpha_2) = \alpha_1$, and $(d_1 \cap d_2, \alpha_1)$ subsumes (d_1, α_1), thus preserving the nestedness of focal elements. In case of conflict, the result is no longer expressed by a possibility distribution (since the focal elements d_1 and d_2 are then disjoint). The disjunctive rule works similarly: it assigns the mass $\min(\alpha_1, \alpha_2)$ to $d_1 \cup d_2$, and 1 to DOM. These rules easily generalize to any number of sources. Then, if $K_1, K_2, ..., K_k$ are the maximal consistent subsets of sources supporting a result d (a disjunction of intervals), then the reliability of the result can be computed by applying these rules.

This calculation can help the user select a proper subset of sources bringing information that is as useful and reliable as possible.

VI. A REAL-WORLD APPLICATION IN AGRONOMY

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 properties, the period of application, and the characteristic features of the field? The risk level for a pesticide to reach surface water is measured by the indicator I_{sur} in [13]. Agronomists try to make a diagnosis based on the value of I_{sur}. Pesticide features depend on the chemical properties of the product while pesticide period application and field properties depend on domain knowledge [14]. This knowledge lies in information sources among which books, databases, and expert knowledge in agronomy. Then values for some features vary w.r.t. sources.

Here, we are interested in the use of pesticide *metsulfuron-methyle* and its influence on the surface water. Two characteristic features of *Metsulfuron-methyle* are needed to compute the indicator I_{sur}, namely $DT50$ and *aquatox* (aqu) (more details on these parameters can be found in [13]; they are not crucial for the understanding of this paper). Table IV (simplified data) gives the values of the parameters $DT50$ and aqu according to 11 different information sources, and

	$DT50$	aqu	α_i^a	β_i^a
BUS	120	?	0.2	
PM10	[7,35]	?	0.7	
PM11	[7,35]	100	0.7	0.9
Com96	[14,180]	0.01	0.5	0.6
Com98	[4,71]	?	0.5	
ARSl	[14,38]	?	0.2	
ARSf	[8,105]	?	0.2	
RIVM	[27,34]	100	0.4	0.3
BUK	[7,190]	?	0.4	
AGXf	[4,100]	0.045	0.8	0.8
AGXl	[20,51]	0.045	0.7	0.8

^{a}α_i and β_i denote the certainty degrees of the source i regarding the parameter $DT50$ and aqu respectively.

Table IV
FEATURES OF *Metsulfuron Methyle*

their certainty degrees for each characteristic. The symbol "?" represents the case when the information source does not give data for the parameter. For this case, the certainty degree is unknown and it will not be taken into account in the computation of the reliability. Agronomists look to find a suitable value for each parameter to be considered for computing the I_{sur} indicator, hence facing an information fusion problem.

Lattice construction. The parameters $DT50$ and aqu are independent. No information is available about sources. Then, the appropriate fusion is the method based on the MCS fusion operator. Therefore, Table VI presents the pre-processing table obtained from initial information provided by sources in Table IV. Table V gives subsets of sources producing the MCS of values for each parameter.

K_1	$\{BUS, Com96, BUK\}$
K_2	$\{PM10, PM11, ARSl, ARSf, Com98, RIVM, BUK, AGXf, AGXl\}$
K_3	$\{PM11, RIVM\}$
K_4	$\{Com96\}$
K_5	$\{AGXf, AGXl\}$

Table V
LABEL OF ALL MCS

	$DT50$	aqu
K_1	120	$\emptyset$
K_2	[27, 34]	$\emptyset$
K_3	$\emptyset$	0.045
K_4	$\emptyset$	0.01
K_5	$\emptyset$	100

Table VI
TABLE IV PRE-PROCESSED

Lattice interpretation. The resulting lattice is given in Figure 4 with 21 values allowing to compute the indicator I_{sur}. The lower arguments represent MCS values of each parameter. For the other arguments in the lattice, the subset of sources (resp. values) is composed of all sources (resp. values) lying in the sub-sets of sources (resp. values) of its sub-arguments. For
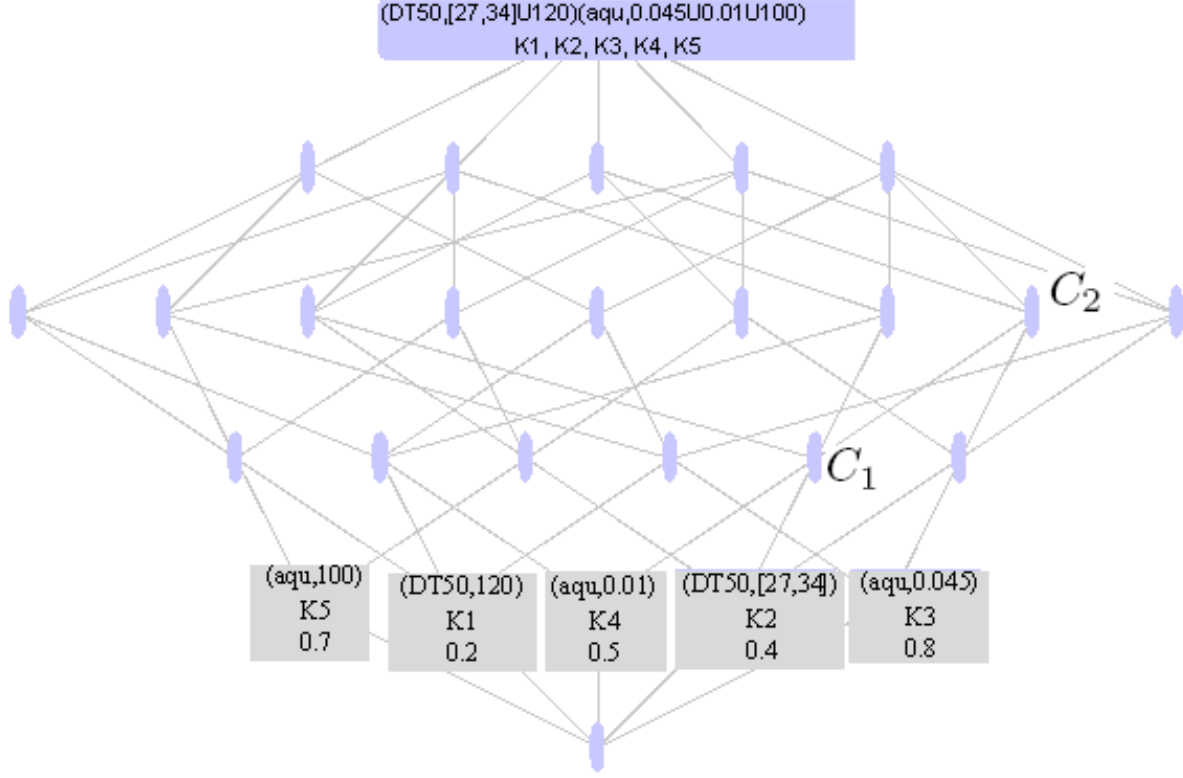
Figure 4. Lattice of MCS fusion results raised from Table IV

example, on Figure 4, the argument value of C_1 is $\{(DT50, [27, 34]), (aqu, 0.01)\}$. But, if two sub-arguments give different values for the same attribute, then the union of values is considered. For example, the argument value of C_2 is $\{(DT50, [27, 34]), (aqu, \{0.045\} \cup \{0.01\})\}$ and its sub-arguments values are $\{(DT50, [27, 34]), (aqu, 0.01)\}$ and $\{(aqu, 0.045)\}$. Moreover, each argument value in the lattice represents the disjunctive fusion result of the subset of sources in the argument. Consequently, these values correspond to the MCS fusion results of subsets in G. Indeed, the highest argument in the lattice corresponds to the MCS global fusion result of all sources for all parameters. Here, this result is $[27, 34] \cup [120, 120]$ for $DT50$ and $\{0.045\} \cup \{0.01\} \cup \{100\}$ for the variable aqu.

First results. Once feature values are summarized from the delivered pieces of information, agronomists try to make a diagnosis w.r.t. the value of I_{sur}. 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 environment-friendly [13].

Considering the global result, i.e. $DT50 = [27, 120]$ and $aqu = [0.01, 100]$ then the indicator $I_{sur} = [3.5, 9.7]^2$. This interval is not useful since some values in $[3.5, 9.7]$

---

are smaller than 7 and other ones are greater than 7. Then, using the set of all sources cannot help agronomists to evaluate the agricultural practices. This due to the diversity of sources and the conflict between sources. Then, it is useful to observe partial fusion results given by the lattice.

For example, considering C_1 (see Figure 4), $DT50 = [27, 34]$ and $aqu = 0.01$, then $I_{sur} = [6.9, 7]$. This indicates that the farmer must change its practices, since values of I_{sur} are smaller than 7. Then, considering the subset $\{PM10, PM11, ARSl, ARSf, Com98, RIVM, BUK, AGXf, AGXl\}$ for $DT50$ and $Com96$ for aqu, using of *metsulfuron methyle* is risky for the environment. However, considering $DT50 = [27, 34]$ and $aqu = 0.045$, the resulting interval of I_{sur} is $[7, 7.4]$ indicating that the practices of the farmer are environment-friendly since the I_{sur} values are greater than 7.

Taking into account the reliability. Consider now that the user knows that some sources are more reliable than others regarding both of parameters $DT50$ and aqu. Each information source is given with a degree of certainty for each parameter (see Table IV). Using the conjunctive combination rule described in section V, we can attach reliability weights to each MCS K_i as shown on figure 4.

---

[2] We use the convex hull of the fusion results for computing I_{sur}

Then it is possible to compute reliability weights for other arguments in the lattice. For instance C_1 merges information about $DT50$ and aqu and has reliability $\min(0.4, 0.6)$. Likewise the argument merging K_1 and K_4 has reliability $\min(0.2, 0.6)$. This computation can help users to identify sources and subset of sources bringing information that is as reliable and useful as possible. Instead of computing a weight for each obtained value of the indicator, handling reliability in the combination of sources using a lattice of arguments allows to rank fusion results for each parameter w.r.t. their usefulness in the agricultural context.

VII. RELATED WORKS AND CONCLUDING REMARKS

The method presented here for handling the reliability of sources is similar to the method proposed by Dubois *et al.* [4] in a logical framework. In this work, authors propose to associate each piece of information not only with its certainty level (as in standard possibilistic logic), but also with the sources which provide it, thus capturing the idea of source-based argument like in our methodology. In this work, each conclusion that can be deduced from such a knowledge base is associated with the subset of sources that supports this conclusion, where each source is itself associated with a certainty level. This approach handles conflicts, by allowing mutually exclusive conclusions supported by distinct subsets of sources. However, conclusions are not merged, as done here with numerical interval values, taking advantage of formal concept analysis for structuring the merging results. Similarly, the work presented here may be also somewhat related with another extension of possibilistic logic called "logic of supporters" [15], where logical formulas are associated with a set of supporters organized in a lattice structure, but this proposal does not deal with information fusion.

In this paper, we propose a method to combine information using formal concept analysis, more particularly its extension for numerical data (pattern structures). This method provides a lattice of arguments. We also detail how to handle the reliability (when available) in the combination process. Instead of giving a unique fusion result, the lattice of arguments gives a structured view of partial fusion results labelled by subsets of sources. Moreover, when the reliability of sources is available, arguments are rank-ordered in order to identify the most appropriate evaluation.

Lastly, it is interesting to notice that the approach presented here in the case of imprecise numerical information represented by means of intervals, could be adapted to any types of uncertainty representations. In particular, these subsets may be the sets of models of propositional knowledge bases. Then, our approach to information fusion would fit in the setting first proposed in [16], and known to be a special case of pattern structure.

REFERENCES

[1] I. Bloch et al., "Fusion: General concepts and characteristics," *Int. J. Intell. Syst.*, vol. 16, no. 10, pp. 1107–1134, 2001.

[2] D. Dubois and H. Prade, "Possibility theory and data fusion in poorly informed environments," *Control Eng. Practice*, vol. 2, pp. 811–823, 1994.

[3] B. Ganter and R. Wille, *Formal cConcept Analysis*. Springer, 1999.

[4] D. Dubois, J. Lang, and H. Prade, "Dealing with multi-source information in possibilistic logic," in *European Conference on Artificial Intelligence*, 1992, pp. 38–42.

[5] B. Ganter and S. O. Kuznetsov, "Pattern Structures and Their Projections," in *Int. Conf. on Conceptual Structures*, ser. LNCS (2120). Springer, 2001, pp. 129–142.

[6] D. Dubois, H. Fargier, and H. Prade, "Multiple source information fusion: a practical inconsistency tolerant approach," in *IPMU, Madrid*, 2000, pp. 1047–1054.

[7] S. Destercke, D. Dubois, and E. Chojnacki, "Possibilistic information fusion using maximal coherent subsets," *IEEE Transactions on Fuzzy Systems*, vol. 17, pp. 79–92, 2009.

[8] Z. Assaghir, M. Kaytoue, A. Napoli, and H. Prade, "Managing information fusion with formal concept analysis," in *MDAI*, ser. LNCS, vol. 6408. Springer, 2010, pp. 104–115.

[9] M. Kaytoue, S. O. Kuznetsov, A. Napoli, and S. Duplessis, "Mining gene expression data with pattern structures in formal concept analysis," *Information Sciences*, vol. 181, no. 10, pp. 1989–2001, 2011.

[10] Z. Assaghir, P. Girardin, and A. Napoli, "Fuzzy logic approach to represent and propagate imprecision in agri-environmental indicator assessment," in *IFSA/EUSFLAT Conf., Lisbon*, 2009, pp. 707–712.

[11] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[12] H. Prade and A. Rico, "Possibilistic evidence," in *ECSQARU 2011, Belfast*, ser. LNCS, W. Liu, Ed., vol. 6717, 2011, pp. 713–724.

[13] H. Van der Werf and C. Zimmer, "An indicator of pesticide environmental impact based on a fuzzy expert system," *Chemosphere*, vol. 36, no. 10, pp. 2225–2249, 1998.

[14] C. Bockstaller, P. Girardin, and H. Van Der Werf, "Use of agro-ecological indicators for the evaluation of farming systems," *European Journal of Agronomy*, vol. 7, no. 1-3, 1997.

[15] C. Lafage, J. Lang, and R. Sabbadin, "A logic of supporters," in *Information, uncertainty and fusion*, B. Bouchon-Meunier, R. Yager, and L. Zadeh, Eds. Kluwer Academic Publishers, 1999, pp. 381–392.

[16] S. Ferré and O. Ridoux, "Searching for objects and properties with logical concept analysis," in *ICCS 2001, Stanford*, ser. LNCS, H. S. Delugach and G. Stumme, Eds., vol. 2120. Springer, 2001, pp. 187–201.