Information Fusion 32 (2016) 12–39

Contents lists available at ScienceDirect

Information Fusion

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

The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks

Didier Dubois[a,b,*], Weiru Liu[b], Jianbing Ma[c], Henri Prade[a]

[a] IRIT, CNRS & Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France
[b] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT7 1NN, UK
[c] School of Computing, Electronics and Maths, Coventry University, Coventry, CV1 5FB, UK

ARTICLE INFO

Article history:
Received 18 September 2015
Revised 22 January 2016
Accepted 10 February 2016
Available online 26 February 2016

Keywords:
Information fusion
Knowledge-based merging
Evidence theory
Combination rules
Plausibility orderings
Possibility theory
Imprecise probability

ABSTRACT

We propose and advocate basic principles for the fusion of incomplete or uncertain information items, that should apply regardless of the formalism adopted for representing pieces of information coming from several sources. This formalism can be based on sets, logic, partial orders, possibility theory, belief functions or imprecise probabilities. We propose a general notion of information item representing incomplete or uncertain information about the values of an entity of interest. It is supposed to rank such values in terms of relative plausibility, and explicitly point out impossible values. Basic issues affecting the results of the fusion process, such as relative information content and consistency of information items, as well as their mutual consistency, are discussed. For each representation setting, we present fusion rules that obey our principles, and compare them to postulates specific to the representation proposed in the past. In the crudest (Boolean) representation setting (using a set of possible values), we show that the understanding of the set in terms of most plausible values, or in terms of non-impossible ones matters for choosing a relevant fusion rule. Especially, in the latter case our principles justify the method of maximal consistent subsets, while the former is related to the fusion of logical bases. Then we consider several formal settings for incomplete or uncertain information items, where our postulates are instantiated: plausibility orderings, qualitative and quantitative possibility distributions, belief functions and convex sets of probabilities. The aim of this paper is to provide a unified picture of fusion rules across various uncertainty representation settings.

1. Introduction

Information fusion is a specific aggregation process which aims to extract truthful knowledge from incomplete or uncertain information coming from various sources [15]. This topic is relevant in many areas: expert opinion fusion in risk analysis [24], image fusion in computer vision [13,14], sensor fusion in robotics [1,61,86], database merging [18,21], target recognition [78], logic [67,68] and so forth. Historically the problem is very old. It lies at the origin of probability theory whose pioneers in the XVIIth century were concerned by merging unreliable testimonies at courts of law [98]. Then, this problem fell into oblivion with the development of statistics in the late XVIIIth century. It was revived in the late XXth century in connection with the widespread use of computers, and the necessity of dealing with large amounts of data coming from different sources, as well as the renewed interest toward process human-originated information, and the construction of autonomous artefacts that sense their environment and reason with uncertain and inconsistent inputs.

Information fusion is often related to the issue of uncertainty modelling. Indeed, sources often provide incomplete or unreliable information, and even if such pieces of information are precise, the fact that they come from several sources often results in conflicts to be solved, as inconsistency threatens in such an environment. The presence of incomplete, unreliable and inconsistent information leads to uncertainty, and the necessity of coping with it, so as make the best of what is available, while discarding the wrong. This is the role of information fusion.

There are many approaches and formats to model information, and several uncertainty theories [51]. The fusion problem in the presence of uncertain or incomplete information has been discussed in each of these settings almost independently of the

* Corresponding author at: IRIT, CNRS and Université de Toulouse, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France.Tel.: +33 561556331.
E-mail address: dubois@irit.fr (D. Dubois).

http://dx.doi.org/10.1016/j.inffus.2016.02.006

other ones [49,80,83,101]. Sometimes, dedicated principles have been stated in order to characterise the specific features of the fusion process in the language of each particular formal setting [69,73,87,109]. Several fusion strategies exist according to the various settings. These strategies share some commonalities but may differ from each other in some aspects due to their specific representation formats (for instance, symbolic vs. numerical).

This paper takes an inclusive view of the current available properties from different theories and investigates the common laws that *must be* followed by these fusion strategies[1]. We argue that some properties are mandatory and some are facultative only. The latter can be useful in certain circumstances, or in order to speed up computation time. It is interesting to notice that although each requested property looks intuitively reasonable on its own, they can be inconsistent when put together. This happens in the problem of merging preferences from several individuals modelled by complete preorderings (Arrow impossibility theorem, see the discussion in [22]). However, the basic mandatory properties of information fusion we propose are globally consistent.

The aim of the paper is to lay bare the specific nature of the information fusion problem. This general analysis yields a better understanding of what fusion is about and how an optimal fusion strategy (operator) can be designed. In particular, information fusion differs from preference aggregation, whose aim is to find a good compromise between several parties. Noticeably, while the result of information fusion should be consistent with what reliable sources bring about, a good compromise in a multiagent choice problem may turn out to be some proposal no party proposed in the first stand. So while they share some properties and methods, we claim that information fusion and preference aggregation do not obey exactly the same principles.

We also wish to show the deep unity of information fusion methods, beyond the particulars of each representation setting. To this aim, we look at special characteristics of each theory and what becomes of fusion principles, what are the fusion rules in agreement with these principles. We will check whether known fusion rules in each theory comply with general postulates of information fusion. We explain how these basic properties can be written in different representation settings ranging from set-based and logic-based representations to possibility theory, belief function theory and imprecise probabilities. These comparisons demonstrate that the proposed basic properties truly reflect the nature of fusion in different settings.

The rest of the paper is organised as follows. The next section presents general features of what can be called an information item. Such features can be extracted from information items in each representation framework. Section 3 presents basic principles of information fusion that apply to information items and discuss their relevance. Some additional and facultative principles are discussed. The problem of merging information is carefully distinguished from the one of preference aggregation. Section 4 instantiates our principles on the crudest representation of an information item, as a set of possible values. When such a set basically excludes impossible values, we show that our setting characterises the method of maximal consistent subsets. The case of merging propositional belief bases, for which a set of postulates, due to Konieczny and Pino-Perez [68], exists, is then discussed. We compare them to our fusion principles, and show that the corresponding Boolean information items in our sense correspond to subsets of most plausible values. The next section discusses the fusion of information items represented by plausibility rankings of possible values, going from ordinal representations to numerical ones in terms of possibility distributions. Again, we compare our instantiated principles with existing proposals, and provide examples of rational fusion rules in our sense. Finally the last section discusses representations that blend set-based and probabilistic formalisms, and account for incomplete information, such as belief functions and imprecise probabilities. We instantiate our principles in each setting, and study the property of known rules for merging belief functions. We also analyse postulates for merging imprecise probabilities proposed by Peter Walley [109] in the light of our general approach.

2. A general setting for representing information items

We call what sources of information provide to an end-user *information items* pertaining to some uncertain entity. An information item is understood as a statement, possibly tainted with uncertainty, forwarded by some source, and describing what the current state of affairs is. In order to define a set of requirements that make sense in different representation settings ranging from logic to imprecise probability, we need to describe several features of an information item, that we consider essential.

Consider a non-empty set of possible worlds or state descriptions or alternatives, one of which is the true one, denoted by $W = \{w_1, \ldots, w_{|W|}\}$ (it will often be the range of some unknown precise entity denoted by x). For simplicity, we restrict ourselves to a finite setting. We assume that there are n agents/sources (sensors, experts, etc.) and the ith one is denoted by e_i. Let T_i denote the information item provided by agent e_i about x. For example T_i can be a set, a probability or a possibility distribution [42], or an ordinal conditional function [104] or a knowledge base.

In this paper, we do not discuss the fusion of precise set-valued entities, such as multisets [17], where sets represent complex entities made of the conjunction of several, possibly identical elements, representing hierarchical data structures [19], or related tuples in relational databases. Such multiset fusion problems can be found when cleaning databases containing duplicate data [18] or for the summarisation of documents. On the contrary, sets used in the representation of uncertain items of information contain mutually exclusive values[2].

Here, an information item indicates which values or states of affairs in W are plausible, and which ones are not, for the uncertain entity or parameter x, according to a source. In that sense an information item is completely attached to the source that supplies it and is not an objective description of the state of affairs. It is a representation of knowledge that is likely to be modified by additional information. An information item T will then be characterised by several features:

- Its *support* $\mathcal{S}(T) \subseteq W$, contains the set of values of x considered not impossible according to information T. Namely, $w \notin \mathcal{S}(T)$ if and only if the value w is considered impossible for the source offering T. One may see $\mathcal{S}(T)$ as a kind of integrity constraint attached to T. If $\mathcal{S}(T) = \emptyset$ then information T is said to be strongly inconsistent. The condition $\mathcal{S}(T) \neq \emptyset$ is a weak form of (internal) consistency.
- Its *core* $\mathcal{C}(T) \subseteq W$, contains the set of values considered fully plausible according to information T. One may see $\mathcal{C}(T)$ as the

[1] Preliminary and partial versions of this paper were presented in two conferences [37,38].

[2] Note that if x is a set-valued attribute, we do not consider the fusion of such precise set-values, e.g. $x = A$. But our approach encompasses the case of incomplete information for set-valued attributes [46]. For instance, if x is the precise time interval when the museum is open, a piece of information like "the museum is open from 9 to 12 h" is imprecise in the sense that what we know from it is that $[9, 12] \subseteq x$. If another source claims that "the museum is open from 14 to 17 h." we may conclude that the museum is open from 9 to 12 h *and* from 14 to 17 h, which here is modelled by $[9, 12] \cup [14, 17] \subseteq x$. However this disjunction is actually obtained by the conjunctive fusion of two sets of time spans, namely $\{A\colon [9, 12] \subseteq A\} \cap \{A\colon [14, 17] \subseteq A\}$.