Combining imprecise probability masses with maximal coherent subsets: application to ensemble classification.

Sébastien Destercke[1] and Violaine Antoine[2]

Abstract

When working with sets of probabilities, basic information fusion operators quickly reach their limits: intersection becomes empty, while union results in a poorly informative model. An attractive means to overcome these limitations is to use maximal coherent subsets (MCS). However, identifying the maximal coherent subsets is generally NP-hard. Previous proposals advocating the use of MCS to merge probability sets have not provided efficient ways to perform this task. In this paper, we propose an efficient approach to do such a merging between imprecise probability masses, a popular model of probability sets, and test it on an ensemble classification problem.

1 Introduction

When multiple sources provide information about the ill-known value of some variable X it is necessary to aggregate these pieces of information into a single model. In the case where the initial uncertainty models are precise probabilities and where the aggregated model is constrained to be precise as well, there are only a few options to combine the information (see [3] for a complete review)

The situation changes when one considers imprecision-tolerant uncertainty theories, such as possibility theory, evidence theory or imprecise probability theory (see [7]). As they extend both set-theoretic and probabilistic approaches[1], these theories can use aggregation operators coming from both frameworks, i.e., they can generalise intersections and unions of sets as well as averaging methods.

Sébastien Destercke

UMR Heudiasyc, Université de Technologie de Compiègne, 60200 Compiègne, France e-mail: sebastien.destercke@hds.utc.fr

Violaine Antoine

LISTIC, Polytech Annecy-Chambéry, 74944 Annecy le Vieux, France e-mail: violaine.antoine@univ-savoie.fr

[1] Except possibility theory, that does not encompass probabilities as special cases.

When there is (strong) conflict between information pieces, both conjunctive (intersection) and disjunctive (union) aggregation face some problems: conjunction results are often empty and disjunction results are often too imprecise to be really useful. A theoretically attractive solution to these problems is to use maximal coherent subsets (MCS) [11], that is to consider subsets of sources who are consistent and that are maximal with this property. Aggregation can then be done by combining conjunction within maximal coherent subsets with other aggregation operators, e.g., disjunction. Practically, the main difficulty that faces this approach is to identify MCS, a NP-hard problem in the general case.

Different solutions have been proposed to combine inconsistent pieces of information within the framework of imprecise probability theory. In [10] and [12], hierarchical models are considered. In [1] and [8], Bayesian-like methods (i.e., using conditional probabilities) of aggregation are proposed. In [9] and [14], non-Bayesian methods are studied (although [14] considers that combination methods should commute with Bayesian updating). In the two latter references, MCS are proposed as a solution to combine information pieces that are partially inconsistent, but no practical methods are given to identify MCS.

In this paper, we concentrate on imprecise probability masses and propose a practical approach to apply MCS inspired combination methods to such models. We work in a non-Bayesian framework. Section 2 recalls the necessary background on imprecise probabilities and information fusion. Section 3 describes our approach, of which the most important part is the algorithm to identify MCS. Finally, Section 4 presents an application to ensemble classification, in which resulting classification models are combined using MCS.

2 Preliminaries

The theory of imprecise probabilities [15] is a highly expressive framework to represent uncertainty. This section presents the basics of imprecise probabilities.

2.1 Imprecise probabilities

Consider a variable X taking values in a finite domain D_x of n elements $\{x_1, x_2, \ldots, x_n\}$. Basically, imprecise probabilities characterize uncertainty about X by a closed convex set $\mathcal{P}$ of probabilities defined on D_x. To this set $\mathcal{P}$ can be associated Lower and upper probabilities that are mappings from the power set 2^{D_x} to $[0, 1]$. They are respectively denoted $\underline{P}$ and $\overline{P}$ and are defined, for an event $A \subseteq D_x$, as $\underline{P}(A) = \inf_{p \in \mathcal{P}} P(A)$ and $\overline{P}(A) = \sup_{p \in \mathcal{P}} P(A)$. These two measures are dual, in the sense that $\underline{P}(A) = 1 - \overline{P}(A^c)$, with A^c the complement of A. Hence, all the information is contained in only one of them.

Alternatively, one can start from a lower measure $\underline{P}$ and compute the convex set $\mathcal{P}_{\underline{P}} = \{P \in \mathbb{P}(D_x) | P(A) \geq \underline{P}(A), \forall A \subseteq D_x\}$ of dominating probability measures ($\mathbb{P}(D_x)$ is the set of all probabilities on D_x). Note that the lower value $P_*(A) = \inf_{P \in \mathcal{P}_{\underline{P}}} P(A)$ need not coincide with $\underline{P}(A)$ in general. If the equality

$P_* = \underline{P}$ holds, then $\underline{P}$ is said to be *coherent*. In this paper, we will deal exclusively with coherent lower probabilities. Note that lower probabilities are not sufficient to represent every possible convex sets of probabilities.To represent any convex set $\mathcal{P}$, one actually needs to consider bounds on expectations (see [15]).

2.2 Imprecise probability masses (IPM)

Usually, the handling of generic sets $\mathcal{P}$ (and sets represented by lower probabilities) represent a heavy computational burden. In practice using simpler models alleviate this computational burden to the cost of a lower expressivity. Imprecise probability masses [4] (IPM) are such simpler models.

IPM can be represented as a family of intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$ verifying $0 \leq l_i \leq u_i \leq 1 \forall i$. The interval bounds are interpreted as probability bounds over singletons. They induce a set $\mathcal{P}_L = \{p \in \mathbb{P}(D_x) | l_i \geq p(x_i) \geq u_i, \forall x_i \in D_x\}$. An extensive study of IPM and their properties can be found in [4].

A set L of IPM is said to be proper if the condition $\sum_{i=1}^n l_i \geq 1 \geq \sum_{i=1}^n u_i$ holds, and $\mathcal{P}_L \neq \emptyset$ if and only if L is proper. Considered sets are always proper, other types having no interest. To guarantee that lower and upper bounds are reachable for each singleton x_i by at least one probability in $\mathcal{P}_L$, the intervals must verify:

$$\sum_{i \neq j} l_j + u_i \leq 1 \text{ and } \sum_{i \neq j} u_j + l_i \geq 1 \quad \forall i. \tag{1}$$

If L is reachable, lower and upper probabilities of $\mathcal{P}_L$ can be computed as follows:

$$\underline{P}(A) = \max(\sum_{x_i \in A} l_i, 1 - \sum_{x_i \notin A} u_i), \quad \overline{P}(A) = \min(\sum_{x_i \in A} u_i, 1 - \sum_{x_i \notin A} l_i). \tag{2}$$

If L is not reachable, a reachable set L' is obtained by applying Eq. (2) to singletons.

2.3 Basic combinations of imprecise probabilities

When M sources provide information, there are three basic ways to combine this information: through a conjunction, a disjunction or a weighted mean. When information is given by credal sets $\mathcal{P}_i$, $i = 1, \ldots, M$, computing these basic combination results present some computational difficulties [9]. Computations become much easier if we consider a set $L_1, \ldots, L_M$ of IPM. In this case, if $l_{i,j}, u_{i,j}$ denote respectively the lower and upper probability bounds on element x_i given by source j, (approximated) combinations are as follows:

- Weighted mean ($L_\sum$): $l_{i,\sum} = \sum_{j=1,M} w_j l_{i,j}$, $\quad u_{i,\sum} = \sum_{j=1,M} w_j u_{i,j}$
- Disjunction ($L_\cup$): $l_{i,\cup} = \min_{j=1,M} l_{i,j}$, $\quad u_{i,\cup} = \max_{j=1,M} u_{i,j}$
- Conjunction ($L_\cap$):

$$l_{i,\cap} = \max(\max_{j=1,M} l_{i,j}, 1 - \sum_{k \neq i} \min_{j=1,M} u_{i,j}), \tag{3}$$

$$u_{i,\cap} = \min(\min_{j=1,M} u_{i,j}, 1 - \sum_{k \neq i} \max_{i=1,M} l_{i,j})$$

In general, the bounds obtained by conjunction (3) may be non-proper, i.e. may result in an empty $\mathcal{P}_{L_\cap}$. $L_1, \ldots, L_M$ have a non-empty intersection iff the following conditions [4] hold:

$$\max_{j=1,M} l_{i,j} \leq \min_{j=1,M} u_{i,j} \text{ for every } i \in [1, n] \tag{4}$$

$$\sum_{i=1,n} \max_{j=1,M} l_{i,j} \leq 1 \leq \sum_{i=1,n} \min_{j=1,M} u_{i,j} \tag{5}$$

The first condition ensures that intervals have a non-empty intersection for every singleton, while the second makes sure that the result is a proper probability interval.

3 Maximal coherent subsets (MCS) and IPM

This section describes the methods to identify and combine MCS.
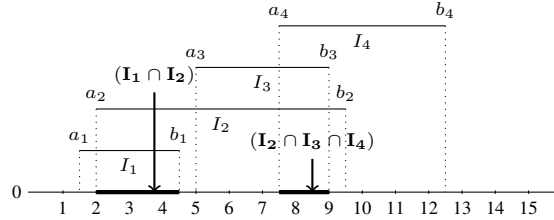
3.1 Identifying MCS

Fig. 1 Maximal coherent subsets on Intervals

When sources provide sets $\mathcal{P}_1, \ldots, \mathcal{P}_M$, finding MCS comes down to find every subset $K \subseteq [1, M]$ such that $\bigcap_{i \in K} \mathcal{P}_i \neq \emptyset$ and such that K is maximum with this property (i.e., adding a new set would make the intersection empty). Usually, identifying every possible coherent subset among $\mathcal{P}_1, \ldots, \mathcal{P}_M$ is NP-hard, making it a difficult problem to solve in practice.

A particularly interesting case where MCS can be found easily is when each sources provide intervals $[a_i, b_i]$, $i = 1, \ldots, M$. In this case, Algorithm 1 given in [6] finds MCS. It requires to sort values $\{a_i, b_i | i = 1, \ldots, M\}$ (complexity in $\mathcal{O}(M \log M)$), and is then linear in the number of sources.

Algorithm 1: Maximal coherent subsets of intervals

Input: M intervals
Output: List of S maximal coherent subsets K_j
1 List = $\emptyset$, j=1, $K = \emptyset$;
2 Order in an increasing order $\{a_i | i = 1, \ldots, M\} \cup \{b_i | i = 1, \ldots, M\}$;
3 Rename them $\{c_i | i = 1, \ldots, 2M\}$ with $type(i) = a$ if $c_i = a_k$ and $type(i) = b$ if $c_i = b_k$;
4 **for** $i = 1, \ldots, 2M - 1$ **do**
5   **if** $type(i) = a$ **then**
6   Add Source k to K s.t. $c_i = a_k$;
7   **if** $type(i + 1) = b$ **then**
8   Add K to List ($K_j = K$) ;
9          $j = j + 1$;
10   **else**
11   Remove Source k from K s.t. $c_i = b_k$;

	$Source1$				$Source2$				$Source3$				$Source4$		
	x_1	x_2	x_3		x_1	x_2	x_3		x_1	x_2	x_3		x_1	x_2	x_3
$u_{i,1}$	0.6	0.5	0.2	$u_{i,2}$	0.55	0.55	0.2	$u_{i,3}$	0.5	0.2	0.6	$u_{i,4}$	0.35	0.6	0.35
$l_{i,1}$	0.4	0.3	0.	$l_{i,2}$	0.35	0.35	0.	$l_{i,3}$	0.3	0.	0.4	$l_{i,4}$	0.15	0.4	0.15

Table 1 Examples of IPM

This algorithm can be applied directly to IPM intervals to check MCS satisfying Condition (4) (which is necessary for a subset of IPM to have a non-empty intersection). Indeed, consider a singleton x_i and the set of intervals $L^i = [l_{i,j}, u_{i,j}]$, $j = 1, M$: if $K \subseteq [1, M]$ is not a MCS of L^i, then the credal sets $\{\mathcal{P}_j | j \in K\}$ do not form a MCS. Hence, iteratively applying Algorithm 1 as exposed in Algorithm 2 allows to easily identify possible MCS among sets $\mathcal{P}_{L_1}, \ldots, \mathcal{P}_{L_M}$. In each iteration (Line 2), Algorithm 2 refines the MCS found in the previous one (stored in $List$) by finding MCS for probaiblity intervals of singleton x_i (Line 5).

Algorithm 2: MCS identification for IPM

Input: M IPM
Output: List of S possible maximal coherent subsets K_j
1 List = $\{\{1, M\}\}$;
2 **for** $i = 1, \ldots, n$ **do**
3   K=$\emptyset$;
4   **foreach** *subset E in List* **do**
5   Run Algorithm 1 on $[l_{i,j}, u_{i,j}], j \in E$;
6   Add resulting list of MCS to K;
7   List= K;

Example 1. Consider the IPM defined on $D_x = \{x_1, x_2, x_3\}$ and summarised in Table 1. Running Algorithm 2 then provides successively the following MCS: $K = \{\{1, 2, 3\}, \{2, 3, 4\}\}$ after the first iteration ($i = 1$ in Line 2 of Algorihm 2); $K =$

$\{\{1, 2\}, \{3\}, \{2, 4\}\}$ after the second iteration, and K is not changed during the third iteration.

Some subsets $K_1, \ldots, K_S$ of sources resulting from Algorithm 2 do not satisfy Condition (5). If K_ℓ is such a set, then one can either make an (exponential) exhaustive search of MCS within K_ℓ, or correct IPM in K_ℓ in a minimal way, so that they satisfy (5). In this last case we can transform [9] bounds $l_{i,j}$ and $u_{i,j}$, $j \in K_\ell, i \in [1, n]$ into $l'_{i,j} = \epsilon l_{i,j}$ and $u'_{i,j} = \epsilon u_{i,j} + (1 - \epsilon)$ with ϵ the minimal value such that

$$\sum_{i=1,n} \max_{j \in K_\ell} l'_{i,j} \leq 1 \leq \sum_{i=1,n} \min_{j \in K_\ell} u'_{i,j}. \tag{6}$$

This strategy makes the identification of MCS easy. Roughly speaking, it applies Algorithm 1 to probabilistic (expectation) bounds coming from different sources but bearing on common events (functions). Note that the same strategy can be applied to models based on peculiar families of events (functions), such as p-boxes [5].

3.2 Combination with MCS

Once MCS $K_1, \ldots, K_S$ of sources have been identified, they can be used to combine inconsistent information. Without loss of generality, consider the indexing such that $|K_1| \geq \ldots \geq |K_S|$ where $|K_i|$ is the cardinality of K_i (i.e., the number of sources within it).

We then propose two ways of combining the probability sets $\mathcal{P}_{L_1}, \ldots, \mathcal{P}_{L_M}$. In both of them, we consider the IPM $L_{K_\ell}, \ell = 1, \ldots, S$ obtained by combining IPM in K_ℓ according to the conjunctive rule (3).

The first rule combines disjunctively the first n IPM L_{K_i}, that is

$$l_{i,\cup\cap_n} = \min_{\ell=1,n} l_{i,K_\ell}, \quad u_{i,\cup\cap_n} = \max_{\ell=1,n} u_{i,K_\ell} \tag{7}$$

where $l_{i,K_\ell}, u_{i,K_\ell}$ are the probability bounds given by L_{K_ℓ} on x_i.

The second rule combines by a weighted mean the first n, L_{K_i}, that is

$$l_{i,\cup\cap_n} = \sum_{\ell=1,n} w_{\ell,n} l_{i,K_\ell}, \quad u_{i,\cup\cap_n} = \sum_{\ell=1,n} w_{\ell,n} u_{i,K_\ell} \tag{8}$$

where $w_{\ell,n} = |K_\ell| / \sum_{i=1,n} |K_i|$ is the importance of K_ℓ in number of sources (a similar strategy is used in [9]). If $n = S$ the rules simply combine every MCS.

4 Application to ensemble classification

Combination is an essential feature of ensemble classification. As classifiers often disagree together, using a MCS based approach to combine the different sources appears sensible. We have therefore tested our approach in the following way: we have trained forest of decision trees; for a given instance and for each decision trees,

we have built an IPM model using the Imprecise Dirichlet model (IDM) with an hyperparameter $s = 4$ (see [2] for details) and taking the samples in the tree leaves as observations. We then combined the different IPM with the two rules (7) and (8) (n=5) and selected the final class according to the maximin and maximality criterion (see [13] for details). The former results in a unique decision while the latter results in a set of possible optimal decisions.

Classifier performances are estimated using discounted accuracy: assume we have T observations whose classes $x^i, i = 1, \ldots, T$ are known and for which T (possibly imprecise) predictions $\widehat{X}^1, \ldots, \widehat{X}^T$ have been made. The discounted accuracy $d - acc$ of the classifier is then

$$d - acc = \frac{1}{T} \sum_{i=1}^{T} \frac{\Delta_i}{f(|\widehat{X}_i|)}, \tag{9}$$

with $\Delta_i = 1$ if $x^i \in \widehat{X}^i$, zero otherwise and f an increasing function such that $f(1) = 1$. Set accuracy $(s - acc)$ is obtained with $f(|\widehat{X}_i|) = 1$.

Results are summarized in Table 2. Numbers of trees in the forest are $\{10, 20, 50\}$ and the data sets are Zoo, Segment and Satimage (taken from UCI), all of them with 7 classes. Results were compared to a classical voting strategy. We have also indicated the average CPU time needed to apply the different combination rules. From the results, it appears that using a conjunctive rule between provided imprecise probabilistic models does not improve much the results of classical voting. This is not surprising as we use precise decision trees and IDM to build our models, and it would be worthwhile to check whether these conclusions still hold when using credal classifiers. The interest of using imprecise probabilistic models appears when we allow for some imprecision, that is when we adopt a partially disjunctive rule (Rule (7) with n=5). In this latter case, allowing for imprecise classification increases the percentage of well-recognized instances while not decreasing too much the precision. Finally, we can notice that the average computational time does not increase much when the number of sources increases.

data set	Tree nb	Single tree	Votes	Rule (7) (n=1)			Rule (7) (n=5)			Rule (8) (n=5)			avg CPU time
				d-acc	s-acc	acc	d-acc	s-acc	acc	d-acc	s-acc	acc	
Sat	10	0.81	0.88	0.87	0.88	0.87	0.64	0.98	0.81	0.87	0.89	0.87	15.16
	20	0.81	0.89	0.88	0.89	0.88	0.61	0.98	0.82	0.88	0.90	0.88	23.80
	50	0.81	0.89	0.89	0.90	0.89	0.63	0.97	0.85	0.89	0.90	0.89	61.28
Zoo	10	0.91	0.92	0.86	0.86	0.86	0.69	0.96	0.80	0.88	0.92	0.92	0.28
	20	0.91	0.93	0.79	0.79	0.79	0.60	0.99	0.68	0.85	0.91	0.88	0.50
	50	0.91	0.93	0.90	0.92	0.91	0.69	0.96	0.87	0.86	0.88	0.88	2.34
Seg	10	0.93	0.96	0.96	0.96	0.96	0.72	1.00	0.84	0.95	0.97	0.96	5.59
	20	0.93	0.95	0.95	0.95	0.95	0.66	0.99	0.81	0.95	0.96	0.95	8.03
	50	0.93	0.96	0.96	0.96	0.96	0.64	0.98	0.84	0.96	0.96	0.96	17.64

Table 2 Results summary. d-acc: discounted accuracy, s-acc: set accuracy, acc: standard accuracy (with maximin)

5 Conclusion

In this paper, we have proposed an efficient way to find MCS with imprecise probability masses, and have applied it to the combination of multiple classifiers. First results indicate that using a disjunctive approach to combine conjunctively merged MCS may quickly result in poorly informative models, hence it may be safer in general to adopt other strategies (e.g., combining only a limited number of MCS or using a weighted mean).

Note that the algorithms presented here can be applied to other imprecise probabilistic models as well, as long as they are defined by probability bounds bearing on the same events (or by expectation bounds bearing on the same function).

References

1. A. Benavoli and A. Antonucci. An aggregation framework based on coherent lower previsions: Application to zadeh's paradox and sensor networks. *Int. J. Approx. Reasoning*, 51(9):1014–1028, 2010.
2. J. Bernard. An introduction to the imprecise dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-3):123–150, 2005.
3. R. Cooke. *Experts in uncertainty*. Oxford University Press, Oxford, UK, 1991.
4. L. de Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
5. S. Destercke and D. Dubois. *The role of generalised p-boxes in imprecise probability models*, pages 179–188. Number 1. 2009.
6. D. Dubois, H. Fargier, and H. Prade. Multi-source information fusion: a way to cope with incoherences. In Cepadues, editor, *Proc. of French Days on Fuzzy Logic and Applications (LFA)*, pages 123–130, La rochelle, 2000. Cepadues.
7. D. Dubois and H. Prade. *Decision-making Process: Concepts and Methods*, chapter Formal representations of uncertainty, pages 85–156. Wiley, 2009.
8. A. Karlsson, R. Johansson, and S. F. Andler. On the behavior of the robust bayesian combination operator and the significance of discounting. In *ISIPTA'09: Proc. of the Sixth Int. Symp. on Imprecise Probability: Theories and Applications*, pages 259–268, 2009.
9. S. Moral and J. Sagrado. Aggregation of imprecise probabilities. In B. BouchonMeunier, editor, *Aggregation and Fusion of Imperfect Information*, pages 162–188. Physica-Verlag, Heidelberg, 1997.
10. R. Nau. The aggregation of imprecise probabilities. *Journal of Statistical Planning end Inference*, 105:265–282, 2002.
11. N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–219, 1970.
12. M. Troffaes. Generalising the conjunction rule for aggregating conflicting expert opinions. *I. J. of Intelligent Systems*, 21(3):361–380, March 2006.
13. M. Troffaes. Decision making under uncertainty using imprecise probabilities. *Int. J. of Approximate Reasoning*, 45:17–29, 2007.
14. P. Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, 1982.
15. P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.