

Ranking from pairwise comparisons in the belief functions framework

Marie-Hélène Masson and Thierry Denœux

Abstract The problem of deriving a binary relation over alternatives based on paired comparisons is studied. The problem is tackled in the framework of belief functions, which is well-suited to model and manipulate partial and uncertain information. Starting from the work of Trichler and Lockwood [8], the paper proposes a general model of mass allocation and combination, and shows how to practically derive a complete or a partial ranking of the alternatives. A small example is provided as an illustration.

1 Introduction

The aim of the paper is to study the task of constructing a linear order, or a ranking, of n alternatives, based on paired comparisons. Paired experiments consist in presenting two objects to one or several judges and asking them to choose the best alternative among the pair. Each paired comparison is supposed to provide uncertain pieces of evidence on the ranking relation, and the derivation of a linear order is considered as an information fusion problem. Uncertain possibilistic preferences have been already considered e.g. in [2]. In this paper, the problem is tackled in the framework of belief functions.

A first work using belief functions to describe the uncertainty about the comparisons has been proposed in [8]. Unfortunately, this work remains essentially theoretical and gives very few tools for practical applications. Our paper synthesizes their main results in Section 3 and extends them in Section 4 in three ways: a more general model of mass allocation is proposed, a linear programming approach for determining the most plausible ranking is introduced, and a heuristic procedure for choosing only a partial order, starting from the most plausible ranking, is given. A small example in Section 5 illustrates the proposed method. Note that, due to space limitations, basic knowledge on belief functions will be assumed. The reader is referred, in particular, to [3, 6].

Marie-Hélène Masson
UPJV, Heudiasyc, UMR CNRS 6599 BP 20529, 60205 Compiègne, France e-mail: mmas-
son@hds.utc.fr

Thierry Denœux
UTC, Heudiasyc, UMR CNRS 6599 BP 20529, 60205 Compiègne, France e-mail:
thierry.denoeux@hds.utc.fr

2 Basic notions on relations

Let $O = \{o_1, o_2, \dots, o_n\}$ be a set of n alternatives. We recall that a strict order R on O is a binary relation for which the following properties hold for all o_i, o_j and $o_k \in O$:

- if $(o_i, o_j) \in R$ then $(o_j, o_i) \notin R$ (asymmetry);
- if $(o_i, o_j) \in R$ and $(o_j, o_k) \in R$ then $(o_i, o_k) \in R$ (transitivity);

If the order is complete (either $(o_i, o_j) \in R$ or $(o_j, o_i) \in R$), it is a linear (or total) order, otherwise it is a partial order. If $(o_i, o_j) \in R$ or $(o_j, o_i) \in R$ then o_i and o_j are comparable, otherwise they are said incomparable.

A linear order L is called a linear extension of a partial order P if $P \subseteq L$ ($\forall (o_i, o_j) \in P$, then $(o_i, o_j) \in L$). To each partial order can thus be associated the set of its linear extensions. Conversely, any collection C of total orders defines a partial order H as follows: $(o_i, o_j) \in H$ iff (o_i, o_j) belongs to all linear order in C . One then says that H is realized by C . Note that two subsets can realize the same partial order.

Any relation R can be conveniently represented by a directed graph with nodes O . Two nodes o_i and o_j are connected by an arc in the graph if $(o_i, o_j) \in R$.

3 Pairwise comparisons in the framework of belief functions

Combining pairwise comparisons in the framework of belief functions has been already addressed by Trichler and Lockwood [8]. This section follows their presentation and synthesizes the most useful notions. They consider that, for each pair (o_i, o_j) of alternatives in O ($1 \leq i < j \leq n$), an expert expresses its preference between o_i and o_j using a mass function $m^{\Theta_{ij}}$ quantifying the uncertainty in the evaluation. This mass function is defined on the frame of discernement $\Theta_{ij} = \{o_i \succ o_j, o_j \succ o_i\}$: the singleton $o_i \succ o_j$ means that o_i should be ranked before o_j and the singleton $o_j \succ o_i$ that o_j should be ranked first. Trichler and Lockwood propose to use a simple support mass function: the expert chooses one of the singletons with mass α_{ij} and the rest of the mass is allocated to Θ_{ij} . The value α_{ij} is interpreted as the reliability of the choice.

Let ϕ_{ij} denote a focal element of $m^{\Theta_{ij}}$. Each focal element ϕ_{ij} has a graph representation which consists of two nodes, o_i and o_j , with one arc if ϕ_{ij} is a singleton element, and no arc if the focal element is Θ_{ij} .

The problem is to derive from the $n(n-1)/2$ mass functions a ranking of the alternatives. This task may be seen as an information fusion problem and Dempster's rule of combination [3] can be used to this end. Let I denote the set $\{(i, j) \mid 1 \leq i < j \leq n\}$ and let $\Theta(I)$ denote the product space:

$$\Theta(I) = \Theta_{12} \times \Theta_{13} \times \dots \times \Theta_{(n-1)n}$$

$\Theta(I)$ consists of all complete asymmetric relations (or graphs) defined on the set O . Before being combined, pieces of evidence from all pairs have to be expressed on the same frame of discernement, namely the product space $\Theta(I)$. This is achieved

by applying the vacuous extension operation [3, 5] to each $m^{\Theta_{ij}}$. This operation, denoted \uparrow , transfers each mass $m^{\Theta_{ij}}(\phi_{ij})$ to $\phi_{ij} \times \Theta(I - \{(i, j)\})$. The symbol \oplus representing Dempster's rule of combination, the expression of the combination can thus be formally written as:

$$m^{\Theta(I)} = m^{\Theta_{12}\uparrow\Theta(I)} \oplus m^{\Theta_{13}\uparrow\Theta(I)} \oplus \dots \oplus m^{\Theta_{(n-1)n}\uparrow\Theta(I)}, \quad (1)$$

or, using the commonalities:

$$q^{\Theta(I)} = \prod_{(i,j) \in I} q^{\Theta_{ij}\uparrow\Theta(I)}. \quad (2)$$

The focal elements of $m^{\Theta(I)}$ are of the form: $\phi = \phi_{12} \times \phi_{13} \times \dots \times \phi_{(n-1)n}$, where ϕ_{ij} is a focal element of $m^{\Theta_{ij}}$ and the mass resulting from the combination is:

$$m^{\Theta(I)}(\phi) = m^{\Theta_{12}}(\phi_{12})m^{\Theta_{13}}(\phi_{13})\dots m^{\Theta_{(n-1)n}}(\phi_{(n-1)n}). \quad (3)$$

In terms of graph, each focal element ϕ of $m^{\Theta(I)}$ can be represented by a directed graph formed by the union of individual graphs. Since each ϕ_{ij} is equal either to a singleton or to Θ_{ij} , each focal element ϕ is a subset composed of complete asymmetric relations on O , whose graphs contain the arcs of ϕ .

The combination described above allocates masses on various sets of asymmetric relations defined on O . A first objective is to find a linear ordering on O that is the most compatible with the pairwise evaluations. This can be done by imposing conditions on the set in which the solution has to be found. Let \mathcal{L} denote the set of all linear orders defined on O which is a subset of $\Theta(I)$. To impose the nature of the solution, it is proposed in [8] to condition the mass $m^{\Theta(I)}$ with respect to \mathcal{L} :

$$m^{\Theta(I)}[\mathcal{L}] = m^{\Theta(I)} \oplus m_{\mathcal{L}}, \quad (4)$$

with $m_{\mathcal{L}}$ a categorical mass function defined by $m_{\mathcal{L}}(\mathcal{L}) = 1$.

Expressed using the commonalities, the whole combination can be written as:

$$q^{\Theta(I)}[\mathcal{L}] = \frac{1}{1-K} q_{\mathcal{L}} \prod_{(i,j) \in I} q^{\Theta_{ij}\uparrow\Theta(I)}, \quad (5)$$

where K is the conflict resulting from the combination of $m^{\Theta(I)}$ with $m_{\mathcal{L}}$. K can be interpreted as an index of the internal coherence of the evaluations. Its practical computation will be explained when dealing with partial orders.

4 Practical use

We consider in this section a general form of mass allocation defined by:

$$\begin{cases} m^{\Theta_{ij}}(o_i \succ o_j) = \alpha_{ij}, \\ m^{\Theta_{ij}}(o_j \succ o_i) = \beta_{ij}, \\ m^{\Theta_{ij}}(\Theta_{ij}) = 1 - \alpha_{ij} - \beta_{ij}. \end{cases} \quad (6)$$

This mass allocation may come from a single expert who is asked to provide, for each $(i, j) \in I$, the above mass function, or from the combination of the evaluations of several experts. In that case, for each $(i, j) \in I$, several $m_k^{\Theta_{ij}}$ are available and they have to be fused to provide $m^{\Theta_{ij}}$. The choice of the combination rule depends on the hypotheses made on the dependence between the experts. If they can be considered as independent, Dempster's rule should be chosen. Otherwise, the cautious rule [1] may be preferred.

4.1 Most plausible ranking

Let $L \in \mathcal{L}$ be a strict linear ordering on O . L being a singleton of the frame of discernement, one has $q^{\mathcal{L}}(\{L\}) = 1$ and, with the mass allocation (6), one has:

$$\begin{cases} q^{\Theta_{ij} \uparrow \Theta(I)}(\{L\}) = 1 - \beta_{ij} & \text{if } (o_i, o_j) \in L, \\ q^{\Theta_{ij} \uparrow \Theta(I)}(\{L\}) = 1 - \alpha_{ij} & \text{if } (o_j, o_i) \in L. \end{cases} \quad (7)$$

Let us introduce $n(n-1)/2$ binary variables $l_{ij}((i, j) \in I)$ defined by $l_{ij} = 1$ if $(o_i, o_j) \in L$ and 0 otherwise. Using (7) and (5), the commonality, or, equivalently, the plausibility of L can be written as:

$$q^{\Theta(I)}[\mathcal{L}](\{L\}) = \text{pl}^{\Theta(I)}[\mathcal{L}](\{L\}) = \frac{1}{1-K} \prod_{(i,j) \in I} (1 - \beta_{ij})^{l_{ij}} (1 - \alpha_{ij})^{1-l_{ij}}, \quad (8)$$

where K is the conflict resulting from the combination of $m^{\Theta(I)}$ with $m^{\mathcal{L}}$. To find the most plausible ranking of the alternatives, it is not necessary to enumerate all possible linear orderings. We propose to solve the problem using a linear programming approach. Maximizing expression (8) is equivalent to maximize its logarithm so that, omitting the constant term depending on K , the most plausible ranking L can be found as the solution of the following linear program:

$$\max_{l_{ij} \in \{0,1\}} \sum_{(i,j) \in I} l_{ij} \ln \left(\frac{1 - \beta_{ij}}{1 - \alpha_{ij}} \right), \quad (9)$$

subject to:

$$\begin{cases} l_{ij} + l_{jk} - 1 \leq l_{ik}, \quad \forall i < j < k, \\ l_{ik} \leq l_{ij} + l_{jk}, \quad \forall i < j < k. \end{cases} \quad (10)$$

The constraints are used to insure that L belongs to \mathcal{L} : if $l_{ij} = 1$ and $l_{jk} = 1$ then $l_{ik} = 1$. If $l_{ij} = 0$ and $l_{jk} = 0$ then $l_{ik} = 0$.

Remark 1 Note that the general form of mass allocation (6) allows us to take naturally into account tied evaluations. If the comparison between o_i and o_j results in a tie, we let $\alpha_{ij} = \beta_{ij}$. Then, it can be easily seen that the pair (o_j, o_j) does not appear any more in the objective function.

4.2 Plausibility of a partial ranking

In some situations, it may be also interesting to compute the plausibility of a partial order. When working in the set of asymmetric relations as defined by Tritchler and Lockwood, it is not possible to provide an analytical expression. However, some results from [8] make it possible to use simple algorithms based on graph theory to compute the plausibility of any partial order. The following theorem is proved in [8]:

Theorem 1 (Tritchler and Lockwood (1991)) Let K be the conflict between the two mass functions $m^{\Theta(I)}$ and $m^{\mathcal{L}}$.

1. $K = \sum m^{\Theta(I)}(\phi)$ where the summation is over every focal element of $m^{\Theta(I)}$ whose graph contains a cycle;
2. Let H be a partial order realized by a focal element θ_H of $m^{\Theta(I)}[\mathcal{L}]$. Then θ_H is the set of all linear extensions of H (θ_H is the largest subset of \mathcal{L} which realizes H).
3. $m^{\Theta(I)}[\mathcal{L}](\theta_H) = \frac{1}{1-K} \sum m^{\Theta(I)}(\phi)$, where the summation is over every focal element ϕ of $m^{\Theta(I)}$ such that the transitive closure of $G(\phi)$, $G^t(\phi)$, is equal to H .

To compute the plausibility of any partial order, one has to sum the masses associated to all focal elements with a non null intersection with this partial order. The following lemma helps to recognize the focal elements which intersect a given partial order:

Lemma 1 (Tritchler and Lockwood (1991)) Let C_1 and C_2 be two subsets of \mathcal{L} realizing the partial orders H_1 and H_2 . Then $C_1 \cap C_2 \neq \emptyset$ if and only if $H_1 \cup H_2$ is acyclic.

Computing the plausibility of any partial order H is thus achieved by summing the masses of all focal sets $\theta_{H'}$ such that $H' \cup H$ is acyclic. .

Lemma 1 and Theorem 1 allow us to propose two simple procedures (one exact, one approximate) for computing the plausibility of a given partial order. The two procedures are detailed in Algorithms 1 and 2.

4.3 Heuristic search for a partial ranking

If the plausibility of the most plausible ranking is too low, it can be preferable to provide the user with only a partial ranking of the alternatives. The algorithms de-

Algorithm 1 Plausibility of a partial order H

```

1:  $K \leftarrow 0$ 
2:  $\text{pl}(H) \leftarrow 0$ 
3: for all  $\phi = \phi_{12} \times \phi_{13} \times \dots \times \phi_{(n-1)n}$  do
4:   Compute the transitive closure  $G^t(\phi)$  of  $G(\phi)$ 
5:   Compute the mass  $m = m^{\Theta(I)}(\phi)$  by equation (3)
6:   if  $G^t(\phi)$  contains a cycle then
7:      $K = K + m$ 
8:   else if  $G(H) \cup G^t(\phi)$  is acyclic then  $\text{pl}(H) = \text{pl}(H) + m$ 
9:   end if
10: end for
11:  $\text{pl}(H) \leftarrow \frac{1}{1-K} \text{pl}(H)$ 

```

Algorithm 2 Approximate computation by a Monte-Carlo approach

```

1:  $N_{\text{pl}} \leftarrow 0$ 
2:  $N_K \leftarrow 0$ 
3: for  $\text{rep} \leftarrow 1, N$  do
4:    $G \leftarrow \emptyset$ 
5:   for each  $(i, j) \in I$  do
6:     With probability  $m^{\Theta_{ij}}(\phi_{ij})$ , randomly select a focal element  $\phi_{ij}$  from the focal elements of  $m^{\Theta_{ij}}$ 
7:     if  $\phi_{ij}$  is a singleton, add the corresponding arc to  $G$ 
8:   end for
9:   Compute the transitive closure  $G^t$  of  $G$ 
10:  if  $G^t$  contains a cycle then
11:     $N_K = N_K + 1$ 
12:  else if  $G(H) \cup G^t$  is acyclic then  $N_{\text{pl}} = N_{\text{pl}} + 1$ 
13:  end if
14: end for
15:  $\hat{K} = \frac{N_K}{N}$ 
16:  $\hat{\text{pl}}(H) = \frac{1}{1-\hat{K}} \frac{N_{\text{pl}}}{N}$ 

```

scribed in the previous section allow us to compute the plausibility of any partial order. Instead of exploring every possible partial orders, which would be practically intractable, we propose a heuristic procedure based on a principle of hierarchical clustering. We start from the most plausible ranking (see Section 4.1). Then, at the first step of the procedure, we compute the plausibility of every partial orders obtained by removing one pair of adjacent alternatives from the total order relation. The most plausible partial order is retained and the corresponding pair of alternatives is “merged”. The process is repeated until all alternatives have been merged into a single one. It is easy to see that the sequence of plausibility values is monotonically increasing. Finally, a partial order with the desired level of plausibility can be chosen by the user.

5 Example

We illustrate the methods described above using an example inspired from [8]. In a study conducted at the Ontario Cancer Institute, subjects were asked to give their preferences about four scenarios describing ethical dilemmas in health care. The preferences for all six possible scenario pairs were obtained. The experts were also asked to rate the reliability of their evaluations. The preferences of a subject can be represented by a directed graph in which the vertices are the scenarios and the edges represent the relation “is preferred to”. The values on the edges represent the reliability given by the expert. The graphs of the experts are given in Figure 1. The fact that graph 1 (left) contains a cycle (ADB) shows that the evaluations of expert 1 are not fully consistent. There is no cycle in graph 2 (right), but the degrees of belief are weaker than for expert 1. The evaluations of each expert are modelled using the mass allocation expressed by equation (6) with either α_{ij} or β_{ij} equal to zero.

Applied individually to each expert, the procedure for deriving a complete ranking of the alternatives (Section 4.1) gives the ranking $A \succ D \succ B \succ C$ with a plausibility of 0.8070 for the first expert, and the ranking $A \succ C \succ D \succ B$ with a plausibility equal to 1 for the second one. The plausibilities thus reflect the internal coherence of the experts. The evaluations of the experts can also be combined before searching for a complete ranking. The results of the combination using Dempster’s rule of combination can be found in Table 1.

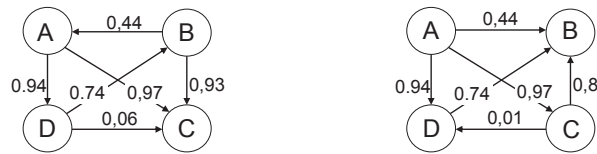


Fig. 1 Graph representation of the evaluations; (left) : expert 1 ; (right) : expert 2.

Table 1 Mass assignment using Dempster’s rule of combination

(o_i, o_j)	$o_i \succ o_j$	$o_j \succ o_i$	Θ_{ij}
(A,B)	0.3056	0.3056	0.3889
(A,C)	0.9991	0	0.0009
(A,D)	0.9964	0	0.0036
(B,C)	0.7266	0.2187	0.0547
(B,D)	0	0.9324	0.0676
(C,D)	0.0594	0.0094	0.9312

The most plausible total ranking derived from Table 1 is $A \succ D \succ B \succ C$ with a plausibility equal to 0.8893. Applying the heuristic procedure using the masses of Table 1 for determining a partial ranking gives the result presented in Figure 2. The

dendrogram can be cut at the desired level of plausibility. For example, the partial order $A \succ D \succ \{B, C\}$ reaches a plausibility of almost 0.96.

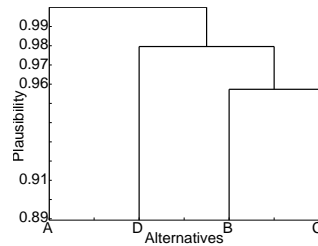


Fig. 2 Dendrogram obtained from Table 1.

6 Conclusion

In this paper, we have shown how to use the framework of belief functions to model paired comparisons and how to derive from these individual judgements a total or a partial ranking of the alternatives. The linear order is obtained by solving a linear program maximizing the plausibility of the relation. A heuristic procedure has been proposed to provide only a partial order when the plausibility of the linear order is too low. This work offers several perspectives, among which the application of the approach to machine learning problems like instance or label ranking problems.

References

1. T. Denoeux. Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence. *Artificial Intelligence*, 172, pp. 234-264 (2008).
2. D. Dubois and H. Prade. On the ranking of ill-known values in possibility theory. *Fuzzy sets and Systems*, 43, pp. 311-317 (1991).
3. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., (1976).
4. P. Smets. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), pp. 447-458 (1990).
5. P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9, pp. 1-35 (1993).
6. P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66, pp. 191-243 (1994).
7. P. Smets. The Transferable Belief Model for quantified belief representation. In D.M. Gabbay and P. Smets, Ed. *Handbook of defeasible reasoning and uncertainty management systems*. volume 1, 267-301, Kluwer Academic Publishers, Dordrecht (1998).
8. D. Trichler and G. Lockwood. Modelling the reliability of paired comparisons. *Journal of Mathematical Psychology*, 35, pp. 277-293 (1991).