AGRI'TERR
AGRO-ÉCOLOGIE DES TERRITOIRES

A formal approach for the representation and the combination of imperfect data

Jérôme Dantan & Salima TAÏBI

Abstract

Nowadays, the sustainability of human activities is a major worldwide concern. Indeed, the problem is no longer to evaluate only the efficiency of human activities, but also sustainability along many axes that can be of various kinds: economic, social, environmental, etc. Such assessments are a major challenge for today's society. Because of the exponential development of means of data recording and storage ("big data" buzz word), and on the basis of Volume, Variety and Velocity properties of big data, scientists need to compute large amounts of data and so do not necessarily have time to clean them. In this context, they compute all available data whose types of imperfections are heterogeneous.

Actors in several domains have to cope with such data, especially to assist humans in their decisions by merging them from many data sources (e.g. measurements, sensors, observations) to model behaviours of complex systems. Mathematical approaches to model imperfect data are well known and established in various scientific areas today, such as both probability based and possibility based calculus. Decisions of experts from various fields have to handle rigorous computations and aggregations of both data and their associated uncertainty. We propose a rigorous model to handle uncertainty on the attributes of objects, and a way to rigorously aggregate discrete data, whose imperfections nature are covered either by the classical probability theory (randomness), either by the possibility theory (fuzziness) thanks to the Dempster-Shafer theory.

Keywords: Probability, Possibility, Dempster Shafer theory

State of the art

Nowadays the scientific problems require reflect uncertainty characteristics, including incompleteness and inaccuracy of information. We introduce here the mathematical modeling approaches and handling of imperfect information relating to possibility theory, probability theory.

Probability theory: probability distributions characterize random phenomena. However, this approach is little suitable for total ignorance representation, and objective interpretations of probabilities, assigned to such events remain difficult when handled knowledge are no longer linked to random and repeatable phenomena (Dubois D., Prade H., 1988). As against, it is possible to model uncertainty thanks to the possibility theory.

Possibility theory: the possibility theory (Dubois D., Prade H., 1988), (Zadeh L.A., 1978) removes the strong probability additive constraint and associates the events of Ω to a possibility measure denoted Π and a necessity measure denoted N, that are both applications from Ω to [0,1], respectively satisfying: $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ and $N(A \cap B) = \min(N(A), N(B))$. This approach also allows representing imprecision using notions of fuzzy sets and distributions of possibilities.

Dempster-Shafer theory: the Dempster-Shafer theory is usually related to neither a probability model nor a possibility model.

In some particular cases, the belief or plausibility deduced from the bbm (basic belief masses) may follow either a probability distribution or a possibility distribution, as stated in the following theorems (Gacôgne L., 1997):

- Theorem 1: the focal elements are totally ordered by inclusion iff Bel and Pl are respectively a measure of necessity and possibility.
- Theorem 2: a belief on a finite set is a probability iff the focal elements are singletons.

Approach

The provided approach provides a formalism for both representing and manipulating rigorously quantities which may have a finite number of possible or probable values.

Let Ω be a universe, with both a possibility measure π and a probability measure P, each having a finite number of values. These values may belong either to a division ring K (e.g. ℝ) or a semigroup G (i.e. a set with an associative internal composition law).

Possibilistic and probabilistic bases

Let E be a vector space over K, infinite but countable dimension. B_I are B and B^J two base sets with:

- $B_I = \{X_{I,i}\ ;\ i = 1, \ldots\}$ (I fixed), we call possibilist bases.
- $B^J = \{X^{J,j}\ ;\ j = 1, \ldots\}$ (J fixed), which we call probabilistic bases.

$D_p(G)$ is the finite set of probability values on G defined on Ω generated by the probabilistic vectors

$D_\pi(G)$ is the finite set of values Possibilists on G defined on Ω generated by the possibilistic vectors

Discrete possibilistic and probabilistic quantities

The canonical form of a purely possibilistic $D_\pi(G)$ a is the following expression: $a = \sum_{i=1}^{n} a_i/\alpha_i.X_{I,i}$ (Dantan et al., 2015), with

- a_i the possible values of a,
- α_i the possibilities associated to each value a_i (one of which at least is equal to 1)
- $X_{I,i}$ (I fixed) the partition of the universe Ω corresponding to values of quantity a

The canonical form of a purely probabilistic $D_p(G)$ b is the following expression: $b = \sum_{i=1}^{n} b_i/\beta_i.X^{J,i}$ (Dantan et al., 2015), with

- b_i the probable values of b,
- β_i are the probabilities, associated to each value bj (the sum of β_i is equal to 1)
- $X^{J,i}$ (J fixed) is the partition of the universe Ω corresponding to values of quantity b.

Combination of possibility and probability distributions

In this section, we present a particular Dempster-Shafer interpretation of combining probabilistic and possibilistic quantites as defined previously. We define the two internal compositions laws + and * on mixed operands:

$$a + b = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(a_i + b_j)/((\alpha_i.X_{I,i}).(\beta_j.X^{J,j}))$$

$$a * b = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(a_i * b_j)/((\alpha_i\,.X_{I,i}).(\beta_j.X^{J,j}))$$

We define φ, as: $\|(\alpha_{1,i}.\ X_{1,i}).(\beta_{2,j}\ .X^{2,j})\| = \varphi(\alpha_{1,i}, \beta_{2,j}).\ X_{1,i}.X^{2,j}$.

φ is a given interpretation of "mixed" possible and probable values.

By hypothesis, we select the interpretation $\varphi(\alpha_{1,i}, \beta_{2,j}) = \alpha_{1,i}\ .\ \beta_{2,j}$

Remark: there may be another composition of $\alpha_{1,i}$, $\beta_{2,j}$, which would be another interpretation of mixing possible and probable values. Indeed, the cross terms $(\alpha_{1,i}.\ X_{1,i}).(\beta_{2,j}\ X^{2,j})$ are assessed at the end of computations.

The couple (Π, P) is associated with a Dempster-Shafer Measure (DSM): this is actually the orthogonal sum of Π and P considered as DSM):

$$F_{i,j} = a_i' \times b_j \text{ where } a_i' = \{a_{\chi(1)}, \ldots, a_{\chi(i)}\}$$

Where χ is a permutation on {1, 2, …, N} such as sequence $a_{\chi(1)}, \ldots, a_{\chi(N)}$ is decreasing, with the associated focal masses:

$$m_M(D_{i,j}) = q_i\ .\ p_j$$

Where q_i is defined by:

- $q_N = \pi_{\chi(N)}$
- $q_i = \pi_{\chi(i)} - \pi_{\chi(i-1)}$ with $N > i \geq 1$ (q_i is positive or null because the χ sequence is increasing)

Use case

Let a and b be respectively a probabilistic and a possibilistic numbers.

- $a = 2\ (probability\ 0.7)\ or\ 3\ (probability\ 0.3)$
- $b = 1\ (possibility\ 1)\ or\ 2\ (possibility\ 0.5)$

a and b are expressed by:

$$a = 2/0.7 + 3/0.3$$

$$b = 1/1 + 2/0.5$$

By applying theorem 1, we may convert the possibilistic distribution b to a bbm distribution, which implies: $m(\{1\}) = 0.5\ ; m(\{1,2\}) = 0.5$

$$a = 2/0.7.\ X^{I,1} + 3/0.3.\ X^{I,2}$$

$$a = 1/1.\ X^{J,1} + 2/0.5.\ X^{J,2}$$

The general case is expressed as follows:

$a + b = 3/0.7.\ X^{I,1}\ .1.\ X^{J,1} + 4/0.7.\ X^{I,2}\ .0.5.X^{J,1} + 4/0.3.\ X^{I,1}\ .1.\ X^{J,2} + 5/0.3.\ X^{I,2}\ .0.5.\ X^{J,2}$

The Dempster-Shafer particular interpretation is expressed as follows:

$a + b = 3/\|0.7.\ X^{I,1}\ .1.\ X^{J,1}\| + 4/\|0.3.\ X^{I,1}\ .1.\ X^{J,2}\| + 4/\|0.7.\ X^{I,2}\ .0.5.X^{J,1}\| + 5/\|0.3.\ X^{I,2}\ .0.5.\ X^{J,2}\|$

$a + b = 3/\varphi(07, 1).\ X^{I,1}\ .\ X^{J,1} + 4/\varphi(0.3, 1)\ X^{I,1}\ .\ X^{J,2} + 4/\varphi(0.7, 0.5).\ X^{I,2}\ .X^{J,1} + 5/\varphi(0.3, 0.5)\ .\ X^{I,2}\ .\ X^{J,2}$

$a + b = 3/0.7.\ X^{I,1}\ .\ X^{J,1} + 4/0.3.\ X^{I,1}\ .\ X^{J,2} + 4/0.35.\ X^{I,2}\ .X^{J,1} + 5/0.15.\ X^{I,2}\ .\ X^{J,2}$

$a + b =$ may have three possible values: 3, 4 or 5 with their respective plausibilities: 0.7, 0.65 (0.3+0.35) et 0.15.

Conclusion

The provided approach provides a formalism for mixing quantities which may have possible or probable values with their interdependencies. The algebraic structure we defined operates on chained computations on such quantities with properties similar to a vector space.

We have extended our formalism on continuous quantities. In special cases such as trapezoids for possibilities and normal distributions for probabilities), some algebraic properties have been maintained. However, combinations of continuous quantities, including probabilistic ones require additional assumptions that imply non rigorous mathematic computations.

The next steps are to compute mixed continuous quantities, by considering either the trapezoidal possibility distributions as intervals of cumulative distribution functions (Destercke S., Dubois D., 2009) or probability-possibility transformations (Dubois et al, 2004).

Dantan, J., Pollet, Y., Taïbi, S. 2015. Combination of imperfect data in fuzzy and probabilistic extension classes. Journal of Environmental Accounting and Management. 3 (2), 123-150. DOI: 10.5890/JEAM.2015.06.004.
Destercke S., Dubois D., 2009. The role of generalised p-boxes in imprecise probability models. In proceedings of 6. International Symposium on Imprecise Probability (p. 179-188). ISIPTA '09, Durham, GBR (2009-07-14 - 2009-07-18).
Dubois D., Foulloy L., Mauris G., Prade H., 2004. Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. Reliable Computing. August 2004, Volume 10, Issue 4, pp 273-297.
Dubois, D. and Prade, H., 1988. Théorie des possibilités, Application à la représentation des connaissances en informatique. Masson 1988. (In French).
Gacôgne L., 1997. Éléments de logique floue. CNAM, Institut d'informatique d'Entreprise, p. 47, may 1997.
Zadeh L.A., 1978. Fuzzy Sets as a basis for a Theory of Possibility. Fuzzy Sets and Systems, 1, 1978.