Fonctions de croyance induites par des ensembles aléatoires flous

Un cadre général pour la représentation des informations floues et incertaines

Thierry Denœux

Université de technologie de Compiègne, CNRS,
UMR 7253 Heudiasyc, France
et
Institut universitaire de France, Paris, France

https://www.hds.utc.fr/~tdenoeux

LFA 2021
Paris, 22 octobre 2021

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 424 (2021) 63–91

FUZZY sets and systems

www.elsevier.com/locate/fss

Belief functions induced by random fuzzy sets: A general framework for representing uncertain and fuzzy evidence

Thierry Denœux [a,b,c,*]

[a] *Université de technologie de Compiègne, CNRS UMR 7253 Heudiasyc, Compiègne, France*
[b] *Institut universitaire de France, Paris, France*
[c] *Shanghai University, UTSEUS, Shanghai, China*

Received 1 April 2020; received in revised form 17 October 2020; accepted 1 December 2020
Available online 9 December 2020

Abstract

We revisit Zadeh's notion of "evidence of the second kind" and show that it provides the foundation for a general theory of epistemic random fuzzy sets, which generalizes both the Dempster-Shafer theory of belief functions and possibility theory. In this perspective, Dempster-Shafer theory deals with belief functions generated by random sets, while possibility theory deals with belief functions induced by fuzzy sets. The more general theory allows us to represent and combine evidence that is both uncertain and fuzzy. We demonstrate the application of this formalism to statistical inference, and show that it makes it possible to reconcile the possibilistic interpretation of likelihood with Bayesian inference.

Keywords: Dempster-Shafer theory; Evidence theory; Possibility theory; Fuzzy mass functions; Uncertain reasoning; Likelihood; Estimation; Prediction

Motivation : inférence statistique

- Soit X un vecteur aléatoire observé, de fonction de densité (ou de masse) de probabilité $f(x, \theta)$, où $\theta \in \Theta$ est un paramètre inconnu.
- Question : Ayant observé $X = x$, comment représenter la connaissance sur le paramètre θ par une fonction de croyance $Bel(\cdot; x)$ sur Θ ?
- Plusieurs solutions. L'une d'elles (proposée pour la première fois par Shafer, 1976) est la fonction de croyance consonante définie par la fonction de plausibilité (mesure de possibilité) :

$$Pl(A) = \sup_{\theta \in A} \widetilde{L}_x(\theta),$$

où $\widetilde{L}_x$ est la fonction de vraisemblance relative (normalisée) définie par

$$\widetilde{L}_x(\theta) = \frac{f(x; \theta)}{\sup_{\theta \in \Theta} f(x; \theta)}$$

($\widetilde{L}_x$ est une distribution de possibilité).

Justification

- Le choix de cette fonction de croyance est intuitivement satisfaisant mais il n'avait pas été justifié par Shafer.
- Dans un article [1] publié en 2014, j'ai montré que cette fonction de croyance est la moins engagée (au sens d'un certain critère) parmi celle vérifiant deux axiomes :
  1. $Bel(\cdot; x)$ doit dépendre uniquement de la fonction de vraisemblance (principe de vraisemblance)
  2. Si P est une mesure de probabilité a priori sur Θ, alors $Bel(\cdot; x) \oplus P = P(\cdot \mid x)$, la mesure de probabilité a posteriori (compatibilité avec l'inférence bayésienne)

1. T. Denœux. Likelihood-based belief function : justification and some extensions to low-quality data. *IJAR*, 55(7) :1535–1547, 2014.

Problème : incompatibilité avec la règle de Dempster

- L'inconvénient de la solution précédente est qu'elle n'est pas compatible avec la règle de Dempster : si x et x' sont des réalisations de deux échantillons indépendants X et X', alors

$$Bel(\cdot; x, x') \neq Bel(\cdot; x) \oplus Bel(\cdot; x')$$

(La fonction de croyance du membre de droite n'est pas consonante).

- En revanche, les distribution de possibilité $\widetilde{L}_x$ et $\widetilde{L}_{x'}$ se combinent par la règle conjonctive de la théorie des possibilité basée sur le produit normalisé $\odot$:

$$\widetilde{L}_{x,x'}(\theta) = (\widetilde{L}_x \odot \widetilde{L}_{x'})(\theta) = \frac{\widetilde{L}_x(\theta)\widetilde{L}_{x'}(\theta)}{\sup_{\theta' \in \Theta} \widetilde{L}_x(\theta')\widetilde{L}_{x'}(\theta')}$$

- Peut-être le cadre de la théorie des possibilités est-il plus approprié que celui de la théorie des fonctions de croyance ?

Prévision statistique

- Cependant, le cadre de la théorie des possibilités semble insuffisant pour représenter l'incertitude dans les problèmes de prévision statistique, dans lesquels entre une composante aléatoire.
- En effet, soit Y une variable aléatoire dont la loi $p(y;\theta)$ dépend du même paramètre $\theta \in \Theta$. On peut toujours écrire [2]

$$Y = \varphi(\theta, U),$$

  où U est une variable aléatoire de loi connue.
- Si l'incertitude (épistémique) sur θ est décrite par la distribution de possibilité $\widetilde{L}_x$, il faut pouvoir la combiner avec l'incertitude aléatoire décrite par la variable aléatoire U, ce qui semble impossible en se plaçant dans le seul cadre de la théorie des possibilités.

2. O. Kanjanatarakul, T. Denœux and S. Sriboonchitta. Prediction of future observations using belief functions : a likelihood-based approach. *IJAR*, 72 :71–94, 2016.

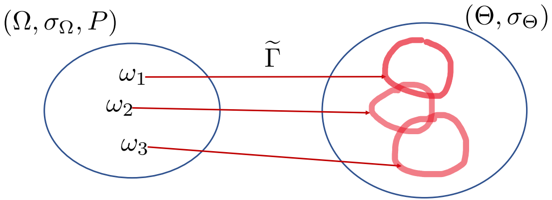
Ensemble flou aléatoire

- Conditionnellement à $U = u$, le principe d'extension de Zadeh permet d'obtenir la distribution de possibilité $\widetilde{Y}_{x,u}$ sur Y à partir de celle sur θ :

$$\widetilde{Y}_{x,u}(y) = \sup_{\{\theta : y = \varphi(\theta,u)\}} \widetilde{L}_c(\theta)$$

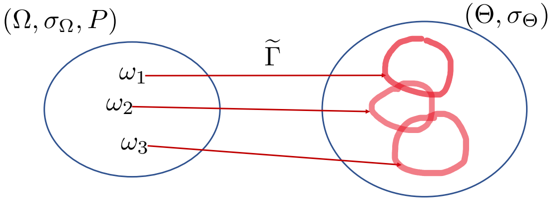
- L'application $u \mapsto \widetilde{Y}_{x,u}$ est un ensemble flou aléatoire (variable aléatoire floue).
- Le cadre de la théorie des ensembles flous aléatoires permet de représenter conjointement et de combiner des informations possibilistes (vraisemblance) et des variables aléatoires.

Définition générale



- Soit $(\Omega, \sigma_\Omega, P)$ un espace probabilisé et (Θ, σ_Θ) un espace probabilisable.
- Une application $\widetilde{\Gamma}$ de Ω dans l'ensemble $\mathcal{F}^*(\Theta)$ des sous-ensembles flous normalisés de Θ est un ensemble flou aléatoire si, pour tout $\alpha \in [0, 1]$, l'application ${}^\alpha\widetilde{\Gamma} : \omega \to {}^\alpha\widetilde{\Gamma}(\omega)$ est un ensemble aléatoire.

Interprétation



- Θ est l'ensemble des valeurs possibles d'une variable Y.
- Ω est une ensemble d'interprétations d'un élément d'évidence.
- Si l'interprétation ω s'applique, alors la valeur de Y est contrainte par la distribution de possibilité $\widetilde{F}(\omega)$.

Cas particulier où Ω est fini

- Lorsque Ω est fini, l'application $\widetilde{\Gamma}$ a un ensemble fini d'images $\widetilde{F}_1, \ldots, \widetilde{F}_f$. On peut définir une application $\widetilde{m} : \mathcal{F}^*(\Theta) \mapsto [0,1]$, appelée fonction de masse floue, telle que

$$\widetilde{m}(\widetilde{F}) = P(\{\omega \in \Omega : \widetilde{\Gamma}(\omega) = \widetilde{F}\}).$$

- Les sous-ensembles flous $\widetilde{F}_1, \ldots, \widetilde{F}_f$ tels que $\widetilde{m}(\widetilde{F}_i) > 0$ sont appelés les ensembles (éléments) focaux de $\widetilde{m}$. On note $\mathbb{F}(\widetilde{m}) = \{\widetilde{F}_1, \ldots, \widetilde{F}_f\}$ et $m_i = \widetilde{m}(\widetilde{F}_i)$.
- Propriété :

$$\sum_{i=1}^{f} m_i = 1.$$

- Dans l'article, Θ est également supposé fini.

Possibilité et nécessité espérées

- Conditionnellement à ω, Y est contrainte par la distribution de possibilité $\widetilde{\Gamma}(\omega)$. Pour tout $A \subseteq \Theta$, les possibilité et nécessité conditionnelles de A sachant ω sont alors

$$\Pi_{\widetilde{\Gamma}(\omega)}(A) = \max_{\theta \in A} \widetilde{\Gamma}(\omega)(\theta)$$

$$N_{\widetilde{\Gamma}(\omega)}(A) = \min_{\theta \notin A}[1 - \widetilde{\Gamma}(\omega)(\theta)] = 1 - \Pi_{\widetilde{\Gamma}(\omega)}(A^c).$$

- On peut alors définir des applications qui associent à tout A sa possibilité et sa nécessité espérées[3] :

$$E\Pi_{\widetilde{m}}(A) = \sum_{\omega \in \Omega} P(\{\omega\}) \Pi_{\widetilde{\Gamma}(\omega)}(A) = \sum_{i=1}^{f} m_i \Pi_{\widetilde{F}_i}(A) = \sum_{i=1}^{f} m_i \max_{\theta \in A} \widetilde{F}_i(\theta)$$

$$EN_{\widetilde{m}}(A) = \sum_{\omega \in \Omega} P(\{\omega\}) N_{\widetilde{\Gamma}(\omega)}(A) = \sum_{i=1}^{f} m_i N_{\widetilde{F}_i}(A) = \sum_{i=1}^{f} m_i \min_{\theta \notin A}[1 - \widetilde{F}_i(\theta)].$$

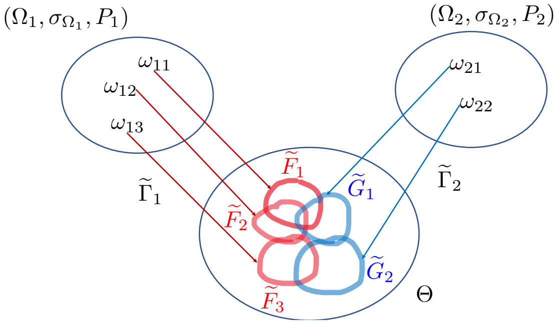
3. L.A. Zadeh. Fuzzy sets and information granularity. In M. M. Gupta et al. eds, *Advances in Fuzzy Sets Theory and Applications*, pp 3–18. North-Holland, Amsterdam, 1979.

Propriétés

- L'application $A \rightarrow EN_{\widetilde{m}}(A)$ est une capacité complètement monotone, c'est-à-dire une fonction de croyance. On la notera $Bel_{\widetilde{m}}$.
- L'application $A \rightarrow E\Pi_{\widetilde{m}}(A)$ est la fonction de plausibilité duale de $Bel_{\widetilde{m}}$. On la notera $Pl_{\widetilde{m}}$.
- Propriété : $Pl_{\widetilde{m}}(A) = 1 - Bel_{\widetilde{m}}(A^c)$ pour tout $A \subseteq \Theta$.
- Remarque importante : à toute fonction de croyance *Bel* correspond une unique fonction de masse « nette »(à éléments focaux nets), mais plusieurs fonctions de masse floues.
- Par exemple, si *Bel* est une mesure de nécessité (i.e., $Bel(A \cap B) = \min[Bel(A), Bel(B)]$ pour tout A et B), il lui correspond
  - Une fonction de masse m nette consonante
  - Une fonction de masse floue $\widetilde{m}$ avec un seul ensemble focal flou défini par

$$\widetilde{F}(\theta) = Pl(\{\theta\})$$

Généralisation de la règle de Dempster



- Soient deux éléments d'évidence représentés par des ensembles flous aléatoires $\widetilde{\Gamma}_1$ and $\widetilde{\Gamma}_2$, supposés indépendants : la mesure de probabilité pertinente sur $\Omega_1 \times \Omega_2$ est la mesure produit $P_1 \otimes P_2$.
- Si les interprétations $\omega_1 \in \Omega_1$ et $\omega_2 \in \Omega_2$ sont vraies, alors la valeur de Y est contrainte par la distribution de possibilité $\widetilde{\Gamma}_1(\omega_1) \cap \widetilde{\Gamma}_2(\omega_2)$, où $\cap$ est un opérateur d'intersection floue normalisée.

Généralisation de la règle de Dempster (suite)

Pour construire un opérateur de combinaison conjonctive d'ensembles flous aléatoires (et donc de fonctions de masse floues dans le cas où Ω est fini), il faut définir :

1. Un opérateur d'intersection normalisée. L'opérateur $\odot$ basé sur la t-norme produit a l'avantage d'être associatif.
2. Une méthode de normalisation : dans la construction de la règle de Dempster, on conditionne la mesure produit $P_1 \otimes P_2$ par l'ensemble des paires d'interprétations (ω_1, ω_2) non contradictoires (dont les images ne sont pas disjointes). Ici, on peut définir un degré de non-contradiction par la hauteur de l'intersection $\widetilde{\Gamma}_1(\omega_1) \cdot \widetilde{\Gamma}_2(\omega_2)$, et conditionner $P_1 \oplus P_2$ par l'ensemble flou des paires non-contradictoires :

$$\Theta^*(\omega_1, \omega_2) = h\left[\widetilde{\Gamma}_1(\omega_1) \cdot \widetilde{\Gamma}_2(\omega_2)\right]$$

Expression de la règle de Dempster généralisée

On obtient finalement l'expression de la règle de Dempster généralisée pour la combinaison de fonctions de masse floues :

$$(\widetilde{m}_1 \oplus \widetilde{m}_2)(\widetilde{F}) := \frac{\sum_{\widetilde{G} \odot \widetilde{H} = \widetilde{F}} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})}{\sum_{(\widetilde{G}, \widetilde{H}) \in \mathbb{F}(\widetilde{m}_1) \times \mathbb{F}(\widetilde{m}_2)} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H})},$$

pour tout $\widetilde{F} \in \mathcal{F}^*(\Theta)$.

Propriétés I

1. Généralisation de la règle de Dempster.
2. Généralisation de la combinaison conjonctive de distributions de possibilité basée sur le produit avec normalisation.
3. Commutativité, associativité.
4. Soient $\widetilde{m}_1$ and $\widetilde{m}_2$ deux fonctions de masse floues de fonctions de contour $pl_{\widetilde{m}_1}$ and $pl_{\widetilde{m}_2}$. La fonction de contour associée à $\widetilde{m}_1 \oplus \widetilde{m}_2$ est

$$pl_{\widetilde{m}_1 \oplus \widetilde{m}_2} = \frac{pl_{\widetilde{m}_1} pl_{\widetilde{m}_2}}{1-\kappa},$$

où κ est le degré de conflit entre $\widetilde{m}_1$ and $\widetilde{m}_2$:

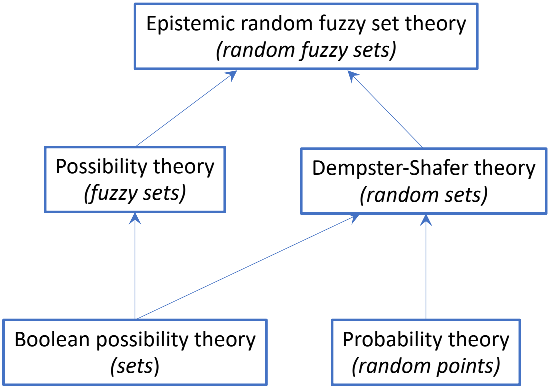
$$\kappa = 1 - \sum_{(\widetilde{G},\widetilde{H}) \in \mathbb{F}(\widetilde{m}_1) \times \mathbb{F}(\widetilde{m}_2)} h(\widetilde{G} \cdot \widetilde{H}) \widetilde{m}_1(\widetilde{G}) \widetilde{m}_2(\widetilde{H}).$$

Propriétés II

5. Soit m une fonction de masse bayésienne correspondant à la mesure de probabilité P et $\widetilde{m}$ une fonction de masse floue logique d'ensemble focal $\widetilde{A} \in \mathcal{F}(\Theta)$. La fonction de masse $m \oplus \widetilde{m}$ est bayésienne et la fonction de croyance correspondante $Bel_{m\oplus\widetilde{m}}$ est égale à la mesure de probabilité $P(\cdot \mid \widetilde{A})$ obtenue en conditionnant P par l'événement flou $\widetilde{A}$: pour tout $A \subseteq \Theta$,

$$Bel_{m\oplus\widetilde{m}}(A) = \frac{\sum_{\theta\in A} P(\{\theta\})\widetilde{A}(\theta)}{\sum_{\theta'\in\Theta} P(\{\theta'\})\widetilde{A}(\theta')} = P(A \mid \widetilde{A}).$$

Résumé



Epistemic random fuzzy set theory
(random fuzzy sets)
Possibility theory
(fuzzy sets)
Dempster-Shafer theory
(random sets)
Boolean possibility theory
(sets)
Probability theory
(random points)

Remarques et perspectives

- Les fonctions de masses floues (*fuzzy belief structures*) ont été étudiées dans les années 1980 et 1990 (Yager, Dubois et Prade, Yen, etc.) et ont été peu utilisées depuis.
- Nous avons revisité cette notion en insistant sur le fait qu'une théorie des « ensembles flous aléatoires épistémiques » peut être vue comme une généralisation de la théorie de Dempster-Shafer et de celle des possibilités.
- Dans article en cours de rédaction,
  - La définition de l'opérateur de Dempster généralisé est étendue à des ensembles flous aléatoires sur des ensembles quelconques (par ex., $\mathbb{R}^n$)
  - Des familles paramétrées d'ensembles flous aléatoires stables par combinaison, permettant de définir des fonctions de croyance sur $\mathbb{R}^n$ facilement manipulables sont introduites.
- Des applications de ces nouveaux modèles en apprentissage automatique sont envisagées.