The Generalization of the Conjunctive Rule for Aggregating Contradictory Sources of Information Based on Generalized Credal Sets

Andrey G. Bronevich[1],
Igor N. Rozenberg[2]

[1] National Research University "Higher School of Economics", Moscow, Russia

[2] JSC Research, Development and Planning Institute for Railway Information Technology, Automation and Telecommunication, Moscow, Russia

9th International Symposium on Imprecise Probability: Theories and Applications,
20-24 July 2015, Pescara, Italy

The conjunctive rule for credal sets

Let X be a finite set and 2^X be the powerset of its subsets.

Definition

A family $\mathbf{P}$ of probability measures on 2^X is called a *credal set* if it is convex and closed.

The *conjunctive rule* (C-rule) for credal sets $\mathbf{P}_1, ..., \mathbf{P}_n$ is defined as

$$\mathbf{P} = \mathbf{P}_1 \cap ... \cap \mathbf{P}_n \tag{1}$$

Remark

The C-rule is defined only for the case when $\mathbf{P}$ is a non-empty set.

The objective of the investigation

1. To extend the C-rule for the case when we have contradictory sources of information, i.e. when the intersection of credal sets may be empty.
2. To construct new models of imprecise probabilities that can work with contradictory information, i.e. when avoiding sure loss condition is not satisfied.

Remark

Observe that we have the analog of the C-rule in the theory of evidence, that can work with contradictory sources of information, and in recent works it has been shown the connection of this C-rule to (1).

The C-rule for belief functions

Let $Bel : 2^X \to [0, 1]$ be a belief function on 2^X, i.e. there is a set function $m : 2^X \to [0, 1]$ called the basic belief assignment (bba) with the following property: $\sum_{A \in 2^X} m(A) = 1$.

If $m(\emptyset) = Bel(\emptyset) > 0$, then Bel describes contradictory information.

Let Bel_i, $i = 1, 2$, be belief functions with bbas m_i. Then the C-rule for Bel_i is defined through the function $m : 2^X \times 2^X \to [0, 1]$ such that

$$\begin{cases} \sum\limits_{A \in 2^X} m(A, B) = m_2(B), \\ \sum\limits_{B \in 2^X} m(A, B) = m_1(B), \end{cases}$$

The C-rule for belief functions

The result of the C-rule has the bba

$$m(C) = \sum_{A \cap B = C} m(A, B).$$

We get the classical C-rule (connected to Dempster's rule) if $m(A, B) = m_1(A) m_2(B)$.
In this case, one can say that sources of information are independent.
This term can be explained through the interpretation of belief functions through random sets.

The choice of optimal C-rules

Obviously there are many C-rules, and there is a problem to choose the optimal one among them. This problem was investigated in
[1] A. G. Bronevich and I. N. Rozenberg. The choice of generalized Dempster-Shafer rules for aggregating belief functions. International Journal of Approximate Reasoning 56: 122-136, 2015.
The main conclusions from [1]:

1. The set of all belief functions is a partially ordered set w.r.t. the specialization order $\preccurlyeq$.
2. Let Bel_i, $i = 1, 2$, be belief functions, then an optimal C-rule should give as a result a belief function Bel being the minimal element of the set

$$\mathbf{Bel}(Bel_1, Bel_2) = \left\{ Bel \in \bar{M}_{bel} | Bel_1 \preccurlyeq Bel, Bel_2 \preccurlyeq Bel \right\},$$

where $\bar{M}_{bel}$ is the set of all belief functions with bbas $m(\emptyset) \geqslant 0$ (it is not necessary that $m(\emptyset) = 0$).

The choice of optimal C-rules

Example 1.

Let P_1 and P_2 be probability measures. Then $\mathbf{Bel}(Bel_1, Bel_2)$ has one minimal element with bba:

$$m(\{x_i\}) = \min\{P_1(\{x_i\}), P_2(\{x_i\})\},$$

$$m(\emptyset) = 1 - \sum_{x_i \in X} \min\{P_1(\{x_i\}), P_2(\{x_i\})\},$$

$$m(A) = 0 \text{ for } |A| \geqslant 2,$$

such belief function can be conceived as a contradictory probability measure if $m(\emptyset) > 0$.

The choice of optimal C-rules

Example 2.

Let Bel_1 and Bel_2 be belief function and

$$Bel(A) = \max\{Bel_1(A), Bel_2(A)\}, \quad A \in 2^X$$

is a belief function such that $Bel_i \preccurlyeq Bel$, $i = 1, 2$.
Then $\mathbf{Bel}(Bel_1, Bel_2)$ has one minimal element $= Bel$.

Example 2 shows the case when the C-rules coincide for belief functions in evidence theory and the theory of imprecise probabilities.

The specialization order

Definition

Let Bel_1 and Bel_2 be belief functions with bbas m_1 and m_2.
$Bel_1 \preccurlyeq Bel_2$ if there is $\Phi : 2^X \times 2^X \to [0,1]$ such that

1) $m_2(B) = \sum\limits_{A \in 2^X} \Phi(A,B) m_1(A)$;

2) $\sum\limits_{B \in 2^X} \Phi(A,B) = 1$, $B \in 2^X$;

3) $\Phi(A,B) = 0$ if $B \not\subseteq A$.

If $Bel_1 \preccurlyeq Bel_2$, then $Bel_1 \leqslant Bel_2$ ($Bel_1(A) \leqslant Bel_2(A)$ for all $A \in 2^X$).
The opposite is not true in general.

Contradiction in information

Let K' be a subset of the set K of all real functions of the type $f : X \to \mathbb{R}$. Then lower previsions on K' are defined by the functional $\underline{E} : K' \to \mathbb{R}$. This functional defines the credal set

$$\mathbf{P}(\underline{E}) = \left\{ P \in M_{pr} | \forall f \in K' : \sum_{x \in X} f(x) P(\{x\}) \geqslant \underline{E}[f] \right\}.$$

If the credal set $\mathbf{P}(\underline{E})$ is empty then lower previsions do not satisfy avoiding sure loss condition and we say that lower previsions contain contradiction.

Contradiction in information

In the same way the upper previsions are defined. The partial case of lower previsions are lower probabilities.
In this case, $K' = \{1_A\}_{A \in 2^X}$ and the functional $\underline{E}$ defines the set function

$$\mu(A) = \underline{E}(1_A),\ A \in 2^X,$$

where we usually assume that μ is a monotone set function with $\mu(\emptyset) = 0$ and $\mu(X) = 1$.
We divide lower probabilities on non-contradictory and non-contradictory ones in the same way.

Problem.

As far as I know in the theory of imprecise probabilities there no models working with contradictory information and this problem appears while we applying the C-rule to the contradictory sources of information.

The expression of the C-rule for models based on lower and upper previsions

Let $\underline{E}_i : K' \to \mathbb{R}$, $i = 1, .., n$, be lower previsions on K'.
Then the result of the C-rule can be expressed as

$$\underline{E}(f) = \max_{i=1,...,n} \underline{E}_i(f),\ f \in K'.$$

Let $\bar{E}_i : K' \to \mathbb{R}$, $i = 1, .., n$, be upper previsions on K'. Then the result of the C-rule can be expressed as

$$\bar{E}(f) = \min_{i=1,...,n} \bar{E}_i(f),\ f \in K'.$$

Clearly, combining non-contradictory information we can get the contradictory information as a result.

The C-rule for probability measures

Case 1.

Probability measures P_1 and P_2 are absolutely contradictory, i.e. there is $A \in 2^X$ such that $P_1(A) = 1$ and $P_2(\bar{A}) = 1$.

$$P_1 \wedge P_2 = \bigwedge_{P_i \in M_{pr}} P_i = \eta^d_{\langle X \rangle}.$$

where the lower probability $\eta^d_{\langle X \rangle}(A) = \begin{cases} 1, & A \neq \emptyset \\ 0, & A = \emptyset, \end{cases}$ describes the result of conjunction of all possible probability measures on 2^X.

This rule can be explained by the law of logic that can be formulated as "*false implies anything*".

The C-rule for probability measures

Case 2.

Probability measures P_1 and P_2 are not absolutely contradictory, then

$$a = 1 - \sum_{x_i \in X} \min\{P_1(\{x_i\}), P_2(\{x_i\})\} < 1,$$

and they can be divided on two parts

$$P_1 = (1-a)P^{(1)} + aP_1^{(2)},\ P_2 = (1-a)P^{(1)} + aP_2^{(2)}$$

such that

$$(1-a)P^{(1)}(\{x_i\}) = \min\left\{P_1(\{x_i\}), P_2(\{x_i\})\right\}.$$

Observe that probability measures $P_1^{(2)}$ and $P_2^{(2)}$ are absolutely contradictory, therefore, the C-rule can be defined as

$$P_1 \wedge P_2 = (1-a)P^{(1)} + a\eta^d_{\langle X \rangle}.$$

The C-rule for probability measures

If we use Dirac measures

$$\eta_{\langle\{x_i\}\rangle}(A) = \begin{cases} 1, & x_i \in A, \\ 0, & x_i \notin A, \end{cases}$$

then the C-rule is expressed as

$$P_1 \wedge P_2 = \sum_{x_i \in X} \min\{P_1(\{x_i\}), P_2(\{x_i\})\}\eta_{\langle\{x_i\}\rangle} + a\eta^d_{\langle X\rangle},$$

where $a = Con(P_1, P_2)$ is the *value of contradiction* between probability measures P_1 and P_2.

The C-rule for probability measures

Example.

Assume that $X = \{x_1, x_2, x_3\}$. Let probability measures P_1 and P_2 be defined by vectors: $P_1 = (0.4, 0.2, 0.4)$, $P_2 = (0.2, 0.4, 0.4)$. Then $a = 0.2$, $P^{(1)}$ =$(0.25, 0, 25, 0.5)$, $P_1^{(2)} = (1, 0, 0)$, $P_2^{(2)} = (0, 1, 0)$.

$$P_1 \wedge P_2 = 0.8 P_1^{(1)} + 0.2 \eta^d_{\langle X \rangle} =$$
$$0.2\eta_{\langle \{x_1\} \rangle} + 0.2\eta_{\langle \{x_2\} \rangle} + 0.4\eta_{\langle \{x_1\} \rangle} + 0.2\eta^d_{\langle X \rangle}.$$

The interpretation of the C-rule through the order $\leqslant$

Notation: $\bar{M}_{cpr}$ is the set of all measures of the type:

$$P = \sum_{i=1}^{n} a_i \eta_{\langle \{x_i\} \rangle} + a_0 \eta^d_{\langle X \rangle} \tag{1}$$

Measures from $\bar{M}_{cpr}$ as lower probabilities describe two types of uncertainty: uncertainty that is essential to probability measures (conflict) and contradiction. If $a_0 = 0$ then obviously formula (1) defines the usual probability measure.

Lemma 1.

Let $P_1 = \sum_{i=1}^{n} a_i \eta_{\langle \{x_i\} \rangle} + a_0 \eta^d_{\langle X \rangle}$, $P_2 = \sum_{i=1}^{n} b_i \eta_{\langle \{x_i\} \rangle} + b_0 \eta^d_{\langle X \rangle}$.
Then $P_1 \leqslant P_2$ (i.e. $P_1(A) \leqslant P_2(A)$) iff $a_i \geqslant b_i$, $i = 1, ..., n$.

The interpretation of the C-rule through the order $\leqslant$

Corollary 1.

Let $P_1, ..., P_m \in \bar{M}_{cpr}$ and $P_k = \sum_{i=1}^{n} a_i^{(k)} \eta_{\langle\{x_i\}\rangle} + a_0^{(k)} \eta_{\langle X\rangle}^d$, $k = 1, ..., m$, then the exact upper bound of $\{P_1, ..., P_m\}$ in $\bar{M}_{cpr}$ w.r.t. the order $\leqslant$ is $P = \sum_{i=1}^{n} c_i \eta_{\langle\{x_i\}\rangle} + c_0 \eta_{\langle X\rangle}^d$, where $c_i = \min\{a_i^{(1)}, ..., a_i^{(m)}\}$, $i = 1, ..., n$, $c_0 = 1 - \sum_{i=1}^{n} c_i$.

Remark.

Corollary 1 implies that the C-rule of probability measures $P_1, P_2 \in M_{pr}$ is the exact upper bound of the set $\{P_1, P_2\}$. Thus, we define next the C-rule for arbitrary measures $P_1, ..., P_m \in \bar{M}_{cpr}$ as the exact bound of the set $\{P_1, ..., P_m\}$ in $\bar{M}_{cpr}$. This bound is denoted as $P_1 \wedge ... \wedge P_m$.

Generalized credal sets

Definition

A subset $\mathbf{P} \subseteq \bar{M}_{cpr}$ is called an *upper generalized credal set* if

1. $P_1 \in \mathbf{P}$, $P_2 \in \bar{M}_{cpr}$, $P_1 \leqslant P_2$ implies that $P_2 \in \mathbf{P}$.
   (The next two properties are essential for the most models of imprecise probabilities (cf. credal sets).)
2. if $P_1, P_2 \in \mathbf{P}$ then $aP_1 + (1-a)P_2 \in \mathbf{P}$ for any $P_1, P_2 \in \mathbf{P}$ and $a \in [0,1]$.
3. the set $\mathbf{P}$ is closed in a sense that it can be considered as a subset of Euclidian space (any $P = a_0\eta^d_{\langle X\rangle} + \sum_{i=1}^{n} a_i\eta_{\langle\{x_i\}\rangle}$ is a vector $(a_0, a_1, ..., a_n)$ in $\mathbb{R}^{n+1}$).

Generalized credal sets

The dual concept to the upper generalized credal set is the lower generalized credal set. Let us remind that

Definition

μ^d is a dual to the monotone measure μ if $\mu^d(A) = 1 - \mu(\bar{A})$.

If we consider measures from $\bar{M}_{cpr}$ as (contradictory) lower probabilities, then any $P^d \in \bar{M}^d_{cpr}$:

$$P = a_0 \eta_{\langle X \rangle} + \sum_{i=1}^{n} a_i \eta_{\langle \{x_i\} \rangle},$$

can be considered as the contradictory upper probability.

Definition

$\mathbf{P}$ is the lower generalized credal set if $\mathbf{P}^d = \left\{ P^d | P \in \mathbf{P} \right\}$ is the upper generalized credal set.

The profile of a generalized credal set

Let $\mathbf{P}$ be a upper generalized credal set in $\bar{M}_{cpr}$. A subset consisting of all minimal elements in $\mathbf{P}$ is called the *profile* of $\mathbf{P}$ and it is denoted by $profile(\mathbf{P})$.

Any profile uniquely defines the corresponding credal set.

If $\mathbf{P}$ describes information without contradiction, then $profile(\mathbf{P})$ is a credal set in usual sense, i.e. $profile(\mathbf{P})$ is a set of probability measures.

Analogously, the profile of lower generalized credal set is defined.

Let $\mathbf{P}$ be a lower generalized credal set in $\bar{M}_{cpr}$. A subset consisting of all maximal elements in $\mathbf{P}$ is called the profile of $\mathbf{P}$ and it is denoted by $profile(\mathbf{P})$.

Obviously, if $\mathbf{P}$ be an upper generalized credal set in $\bar{M}_{cpr}$, then $\mathbf{P}^d$ is the lower generalized credal set in $\bar{M}_{cpr}^d$ and

$$profile(\mathbf{P}^d) = profile(\mathbf{P})^d.$$

The C-rule for generalized credal sets

Definition.

Let $\mathbf{P}_1, ..., \mathbf{P}_m$ be non-empty upper generalized credal sets in $\bar{M}_{cpr}$. Then the credal set $\mathbf{P}$ produced by the C-rule is defined as $\mathbf{P} = \mathbf{P}_1 \cap ... \cap \mathbf{P}_m$.

Let us observe that this definition generalizes the introduced C-rule for probability measures. Actually, let we have two credal sets $\mathbf{P}_1, \mathbf{P}_2$ in $\bar{M}_{cpr}$ with $profile(\mathbf{P}_i) \in M_{pr}$, where M_{pr} is the set of all probability measures on 2^X. Then

$$profile(\mathbf{P}_1 \cap \mathbf{P}_2) = profile(\mathbf{P}_1) \wedge profile(\mathbf{P}_2).$$

In the same way the C-rule for lower generalized credal sets are defined.

The C-rule for generalized credal sets (Example)

Let $X = \{x_1, x_2, x_3\}$. Then any

$$P = a_1\eta_{\langle\{x_1\}\rangle} + a_2\eta_{\langle\{x_2\}\rangle} + a_3\eta_{\langle\{x_3\}\rangle} + a_0\eta^d_{\langle X\rangle}$$

in $\bar{M}_{cpr}$ can be defined by the vector $P = (a_1, a_2, a_3, a_0)$. Consider upper generalized credal sets $\mathbf{P}_i$, $i = 1, 2, 3$, whose profiles are credal sets in usual sense:

$$profile(\mathbf{P}_1) = \{aP_1 + (1-a)P_2 | t \in [0,1]\},\ profile(\mathbf{P}_2) = \{P_3\},$$
$$profile(\mathbf{P}_3) = \{P_4\},$$

where

$$P_1 = (2/3, 0, 1/3, 0),\ P_2 = (0, 2/3, 1/3, 0),\ P_3 = (1/3, 1/3, 1/3, 0),$$
$$P_4 = (1/3, 1/2, 1/6, 0).$$

The C-rule for generalized credal sets (Example)

Let us find the profile of $\mathbf{P}_1 \cap \mathbf{P}_2$. It obviously consists of minimal elements in the set

$$\left\{P' \wedge P'' | P' \in profile(\mathbf{P}_1), P'' \in profile(\mathbf{P}_2)\right\} =$$

$$\{P | P = (1/3, t, 1/3, 1/3 - t), t \in [0, 1/3]\} \cup$$
$$\{P | P = (t, 1/3, 1/3, 1/3 - t), t \in [0, 1/3]\}.$$

The above set has only one minimal element $P_5 = (1/3, 1/3, 1/3, 0)$, therefore, $profile(\mathbf{P}_1 \cap \mathbf{P}_2) = \{P_5\}$.

The C-rule for generalized credal sets (Example)

Analogously, let us find the profile of $\mathbf{P}_1 \cap \mathbf{P}_3$. It consists of minimal elements in the set

$$\left\{P' \wedge P'' | P' \in profile(\mathbf{P}_1), P'' \in profile(\mathbf{P}_3)\right\} =$$

$$\{P|P = (2t/3, 1/2, 1/6, 1/3 - 2t/3), t \in [0, 1/4)\} \cup$$

$$\{P|P = (2t/3, 2(1-t)/3, 1/6, 1/6), t \in [1/4, 1/2]\} \cup$$

$$\{P|P = (1/3, 2(1-t)/3, 1/6, 2t/3 - 1/6), t \in (1/2, 1]\}.$$

The minimal elements of this set are $tP_6 + (1-t)P_7$, where $t \in [0, 1]$, and

$$P_6 = (1/6, 1/2, 1/6, 1/6),\ P_7 = (1/3, 1/3, 1/6, 1/6).$$

Thus, $profile(\mathbf{P}_1 \cap \mathbf{P}_3) = \{tP_6 + (1-t)P_7 | t \in [0, 1]\}$.

The ways of defining generalized credal sets

Let lower probability $P \in \bar{M}_{cpr}$, i.e. $P = a_0 \eta^d_{\langle X \rangle} + \sum_{i=1}^{n} a_i \eta_{\langle \{x_i\} \rangle}$ and $f : X \to \mathbb{R}$.

Then the lower expectation $\underline{E}_P(f)$ of f w.r.t. P can be computed by the Choquet integral:

$$\underline{E}_P(f) = (C) \int f dP = a_0 \max_{x \in X} f(x) + \sum_{i=1}^{n} a_i f(x_i).$$

Let $\underline{E} : K' \to \mathbb{R}$ be a lower prevision. Then it defines the upper generalized credal set

$$\mathbf{P}(\underline{E}) = \left\{ P \in \bar{M}_{cpr} | \forall f \in K' : \underline{E}_P(f) \geqslant \underline{E}(f) \right\}.$$

This set is not empty if $\underline{E}(f) \leqslant \max_{x \in X} f(x)$ for all $f \in K'$.

Upper and lower previsions based on generalized credal sets

We can define the lower prevision based on non-empty upper generalized credal set $\mathbf{P}$ as

$$\underline{E}_{\mathbf{P}}(f) = \inf_{P \in \mathbf{P}} \underline{E}_P(f),\ f \in K.$$

Analogously, we can define the upper prevision based on non-empty lower generalized credal set $\mathbf{P}$ as

$$\bar{E}_{\mathbf{P}}(f) = \sup_{P \in \mathbf{P}} \bar{E}_P(f),\ f \in K.$$

Functionals $\underline{E}_{\mathbf{P}}$ and $\bar{E}_{\mathbf{P}}$ can be considered as counterparts of coherent lower und upper previsions in the theory of imprecise probabilities without contradiction. In the paper, we give necessary and sufficient conditions when a given functional coincides with $\bar{E}_{\mathbf{P}}$ for some credal set $\mathbf{P}$ and describe the construction, which is analogous to natural extension in the traditional theory of imprecise probabilities.

Future work

1. To develop the theory of imprecise probabilities that can work with contradictory information based on generalized credal sets and lower and upper previsions that, maybe, do not avoid sure loss.
2. To consider how the proposed C-rule can be applied to decision problems.

One way is the following. Let we get an upper generalized credal set $\mathbf{P}$ after applying the C-rule. Suppose that the amount of contradiction in $\mathbf{P}$ is

$$Con(\mathbf{P}) = \inf \left\{ Con(P) | P \in \mathbf{P} \right\},$$

where $Con(P) = a_0$ if $P = a_0 \eta^d_{\langle X \rangle} + \sum_{i=1}^{n} a_i \eta_{\langle \{x_i\} \rangle}$.

Future work

Then it is possible to transform contradiction to non-specificity (imprecision), which is used in Yager's rule of combination. In this transformation each

$$P = a_0\eta^d_{\langle X\rangle} + \sum_{i=1}^{n} a_i\eta_{\langle\{x_i\}\rangle}$$

in $\mathbf{P}$ with $a_0 = Con(\mathbf{P})$ is transformed to non-contradictory lower probability

$$P' = a_0\eta_{\langle X\rangle} + \sum_{i=1}^{n} a_i\eta_{\langle\{x_i\}\rangle}.$$

Then it is possible to use well known decision rules developed in the traditional theory of imprecise probabilities.

Thanks for you attention

brone@mail.ru
I.Rozenberg.gismps.ru