

# О выборе источников информации для агрегирования в рамках теории свидетельств

**А.Е. Лепский**

МЦАВР НИУ ВШЭ, Москва

Круглый стол, 2 февраля, 2022, Москва

# Задача агрегирования информации из разных источников

Предположим, что несколько экспертов делают прогноз о стоимости нефти в конце года. Эта информация может быть **интервально-частотной**. Например, {стоимость нефти в конце года будет 78–85 у.е. с достоверностью 0.6, 73–80 у.е. с достоверностью 0.3 или «неизвестно какая»}.

Требуется агрегировать такую информацию **с учетом**:

- различной **надежности** источников информации (экспертов);
- различной **осторожности** экспертов.

**Задачи:**

- что агрегировать?
- как агрегировать?

# Инструментарий теории свидетельств

Пусть

- $X$  – некоторое множество (в нашем случае  $X \subseteq \mathbb{R}$ );
- $\mathcal{A} \subseteq 2^X$  – конечное подмножество **фокальных элементов** (в нашем случае это промежутки вида  $[a, b)$ );
- $m : 2^X \rightarrow [0, 1]$ ,  $\sum_{A \in \mathcal{A}} m(A) = 1$  – **функция масс**,  $m(A) > 0 \forall A \in \mathcal{A}$ ;
- пара  $F = (\mathcal{A}, m)$  называется **телом свидетельств** (свидетельством);
- свидетельство  $F_A = (A, 1)$  называется **категоричным**,  $F_X$  – **бессодержательным**.

Любое свидетельство можно представить в виде  $F = (\mathcal{A}, m)$ :

$$F = \sum_{A \in \mathcal{A}} m(A) F_A.$$

## Правила агрегирования

Например, свидетельство {стоимость нефти в конце года будет 78–85 у.е. с достоверностью 0.6, 73–80 у.е. с достоверностью 0.3 или «неизвестно какая»} запишется так

$$F = 0.6F_{[78,85]} + 0.3F_{[73,80]} + 0.1F_X,$$

где (например)  $X = [10, 110]$ .

Пусть есть два свидетельства  $F_1 = (\mathcal{A}_1, m_1)$  и  $F_2 = (\mathcal{A}_2, m_2)$ .

### Правила агрегирования:

- конъюнктивное правило (ненормализованное):

$$m_{\cap}(C) = \sum_{A \cap B = C} m_1(A)m_2(B).$$

- дизъюнктивное правило:

$$m_{\cup}(C) = \sum_{A \cup B = C} m_1(A)m_2(B).$$

## Конфликт свидетельств

Что агрегировать?  $\Rightarrow$  наименее конфликтные свидетельства

Меры конфликта:

- классическая мера, связанная с правилом Демпстера

$$Con_0(F_1, F_2) = \sum_{A \cap B = \emptyset} m_1(A)m_2(B),$$

- взвешенная классическая мера

$$Con^\Gamma(F_1, F_2) = \sum_{A \in \mathcal{A}_1, B \in \mathcal{A}_2} \gamma(A, B)m_1(A)m_2(B),$$

где  $\Gamma = (\gamma(A, B))_{A, B \in \mathcal{A}}$ ,  $\gamma(A, B) = 1 - s(A, B)$

и  $s(A, B)$  **индекс сходства**, удовлетворяющий условиям:

- $0 \leq s(A, B) \leq 1$ ;
- $s(A, B) = 0$ , if  $A \cap B = \emptyset$ ;
- $s(A, A) = 1$  или (более слабое условие)  $\max_B s(A, B) = s(A, A)$ .

**Примеры.** 1) индекс Жаккара  $s(A, B) = \frac{|A \cap B|}{|A \cup B|}$ ; 2)  $s(A, B) = \frac{|A \cap B|}{|X|}$ .

Мера конфликта  $Con^\Gamma$  будет согласовано с **правилом Цанга** (Zhang)

$$m_{1,2}(C) = \frac{1}{K} \sum_{A \cap B = C} s(A, B) m_1(A) m_2(B),$$

если  $K = 1 - Con^\Gamma(F_1, F_2) \neq 0$ . Обозначение

$$F_{1,2} = (\mathcal{A}_{1,2}, m_{1,2}) = F_1 \otimes F_2.$$

**Надежность** источников информации может быть учтена с помощью дисконтирования – **методом Шафера** (Shafer)

$$m^{(\eta)}(A) = \eta m(A), \quad \text{if } A \neq X$$

и

$$m^{(\eta)}(X) = 1 - \eta + \eta m(X), \quad \eta \in [0, 1].$$

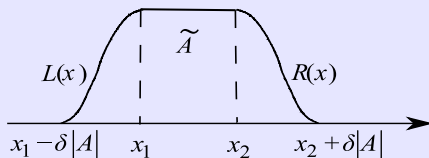
## Преобразование свидетельств

Различные степени надежности и осторожности экспертов можно моделировать с помощью преобразования свидетельств.

Пусть  $\tilde{F} = (\tilde{A}, \tilde{m})$  – преобразование свидетельства  $F = (A, m)$ , где

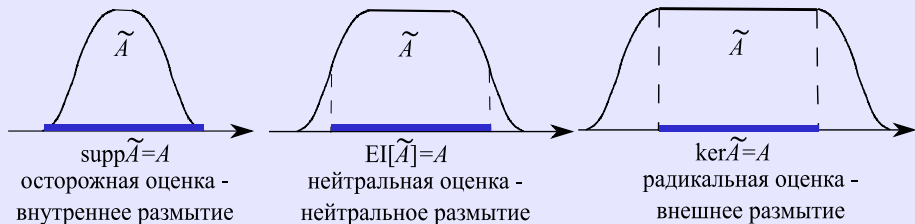
- $\tilde{m}$  – дисконтирование масс;
- $\tilde{A}$  – размытие фокального элемента.

Размытие фокального элемента  $A = [a_1, a_2]$  будем моделировать с помощью нечеткого числа (L-R)-типа  $\tilde{A}$ .



## Стратегии размытия

- если ЛПР даёт часто **осторожные** оценки, то  $\text{supp } \tilde{A} = A$  (внутреннее размытие);
- если ЛПР даёт часто слишком **радикальные** оценки, то  $\text{ker } \tilde{A} = A$  (внешнее размытие);
- если оценки ЛПР **нейтральны**, то  $\text{EI}[\tilde{A}] = \bar{A}$ , где  $\text{EI}$  – ожидаемый интервал нечеткого числа (нейтральное размытие).





## Требование к преобразованиям

### Definition

Меру конфликта  $Con$  назовем:

- устойчивой** к преобразованиям  $\tilde{\phantom{x}}$ , если  $\forall F_1, F_2 \in \mathcal{F}(X)$  и  $\forall \varepsilon > 0$   
 $\exists \delta_1, \delta_2 > 0: \left| Con(\tilde{F}_1, \tilde{F}_2) - Con(F_1, F_2) \right| < \varepsilon$   
 $\forall \tilde{F}_1 = (\tilde{\mathcal{A}}_1, \tilde{m}_1) \in \mathcal{F}(X): d(\tilde{\mathcal{A}}_1, \mathcal{A}_1) = \sum_{A \in \mathcal{A}_1} d(\tilde{A}, A) < \delta_1,$   
 $d(\tilde{m}_1, m_1) = \sum_{A \in \mathcal{A}_1} |\tilde{m}(A) - m(A)| < \delta_2;$
- монотонной** относительно преобразования  $\tilde{\phantom{x}}$ , если  
 $\forall F_1, F_2, F_3 \in \mathcal{F}(X): Con(F_1, F_2) \leq Con(F_1, F_3) \Rightarrow$   
 $Con(\tilde{F}_1, \tilde{F}_2) \leq Con(\tilde{F}_1, \tilde{F}_3).$

### Definition

**Направленность изменений.** Преобразование  $\tilde{\phantom{x}}$  назовем не увеличивающим (не уменьшающим) меру конфликта  $Con$ , если  $Con(\tilde{F}_1, \tilde{F}_2) \leq Con(F_1, F_2)$  ( $Con(\tilde{F}_1, \tilde{F}_2) \geq Con(F_1, F_2)$ )  $\forall F_1, F_2 \in \mathcal{F}(X)$ .

## Численный пример

**Описание данных.** 7 крупнейших инвестбанков в ноябре 2020 г. дали свои прогнозы относительно стоимости нефти в 2021 году. Каждый прогноз – промежуток  $A_i = [a_i, b_i)$ , где  $a_i, b_i$  – прогнозы  $i$ -го инвестбанка относительно стоимости нефти марки Brent в 4-м квартале 2020 и в 4-м квартале 2021 соответственно <sup>1</sup>. Каждый прогноз – это категорическое свидетельство  $F_{A_i}, i = 1, \dots, 7$ .

	investment banks	$a_i$	$b_i$	$\eta_i$	$l_i$
$A_1$	BNP Paribas	45	59	0.8	14
$A_2$	Citigroup	44	56	0.71	12
$A_3$	RBC	41	55	0.63	14
$A_4$	JPMorgan	39	52	0.53	13
$A_5$	Bank of America	47	51	0.67	4
$A_6$	Deutsche Bank	45	50	0.61	5
$A_7$	Standard Chartered	35	50	0.41	15

[www.wsj.com/articles/oil-prices-seen-remaining-subdued-into-2021-11606490833](http://www.wsj.com/articles/oil-prices-seen-remaining-subdued-into-2021-11606490833)

**Надежность**  $\eta_i$  категоричного свидетельство  $F_{A_i}$  определим как убывающую функцию от отклонения середины прогнозного промежутка  $E(F_{A_i}) = \frac{a_i+b_i}{2}$  от текущей (на момент исследования 1.03.2021) цены марки Brent  $c_0 = 65$ :

$$\eta_i = \frac{10 + \max d_k - d_i}{15 + \max d_k - \min d_k},$$

where  $d_i = |E(F_{A_i}) - c_0|$ .

Длина  $l_i = l(A_i) = b_i - a_i$  прогнозного промежутка характеризует степень **неопределенности** прогноза и его осторожность/радикальность.

Промежуток  $X = [20, 80]$  рассматривался как **базовое множество** (нижняя и верхние границы цены за последние три года на момент исследования).

Значения меры конфликта категоричных свидетельств

$$Con^{\Gamma}(F_{A_i}, F_{A_j}) = 1 - s(A_i, A_j)$$

без дисконтирований и размытий.

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$
$A_1$	0	<b>0.27</b>	0.44	0.65	0.71	0.64	0.79
$A_2$	0.27	0	<b>0.27</b>	0.53	0.67	0.58	0.71
$A_3$	0.44	0.27	0	<b>0.31</b>	0.71	0.64	0.55
$A_4$	0.65	0.53	0.31	0	0.69	0.61	<b>0.35</b>
$A_5$	0.71	0.67	0.71	0.69	0	0.5	0.81
$A_6$	0.64	0.58	0.64	0.61	0.5	0	0.67
$A_7$	0.79	0.71	0.55	0.35	0.81	0.67	0

Пары свидетельств для агрегирования в порядке убывания приоритетов в соответствии с принципом минимума конфликта:

$$(F_1, F_2) \sim (F_2, F_3) \succ (F_3, F_4) \succ (F_4, F_7).$$

Мера конфликта

$$\begin{aligned} \text{Con}^\Gamma(\eta_i F_{A_i}, \eta_j F_{A_j}) = & 1 - \eta_i \eta_j s(A_i, A_j) - \\ & - (1 - \eta_i) \eta_j s(X, A_j) - \eta_i (1 - \eta_j) s(A_i, X) - (1 - \eta_i)(1 - \eta_j) \end{aligned}$$

с дисконтированием и без размытий.

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$
$A_1$	0.25	<b>0.44</b>	0.55	0.65	0.71	0.66	0.68
$A_2$	0.44	0.33	<b>0.47</b>	0.59	0.69	0.64	0.63
$A_3$	0.55	0.47	0.36	<b>0.49</b>	0.69	0.64	0.54
$A_4$	0.65	0.59	0.49	0.39	0.68	0.62	<b>0.47</b>
$A_5$	0.71	0.69	0.69	0.68	0.41	0.63	0.7
$A_6$	0.66	0.64	0.64	0.62	0.63	0.44	0.62
$A_7$	0.68	0.63	0.54	0.47	0.7	0.62	0.36

Приоритетные пары для агрегирования:

$$(F_1, F_2) \succ (F_2, F_3) \sim (F_4, F_7) \succ (F_3, F_4).$$

Добавим к дисконтированию масс **смешанное размытие** фокальных элементов: будем использовать **внутреннее** размытие для **'больших'** по длине  $l_i$  фокальных элементов и **внешнее** размытие для **'малых'** фокальных элементов. Фокальные элементы  $A_5$  и  $A_6$  можно считать **'малыми'**, а остальные – **'большими'**. Функция размытия  $\delta(\eta) = 1 - \eta$ .

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$
$A_1$	0.26	<b>0.5</b>	0.64	0.75	0.63	0.69	0.76
$A_2$	0.5	0.35	<b>0.55</b>	0.7	<b>0.56</b>	0.61	0.73
$A_3$	0.64	0.55	0.4	0.59	0.59	<b>0.51</b>	0.65
$A_4$	0.75	0.7	0.59	0.44	0.7	0.58	0.57
$A_5$	0.63	0.56	0.59	0.7	0.4	0.58	0.76
$A_6$	0.69	0.61	0.51	0.58	0.58	0.42	0.66
$A_7$	0.76	0.73	0.65	0.57	0.76	0.66	0.43

Приоритетные пары для агрегирования:

$$(F_1, F_2) \succ (F_3, F_6) \succ (F_2, F_3) \sim (F_2, F_5).$$

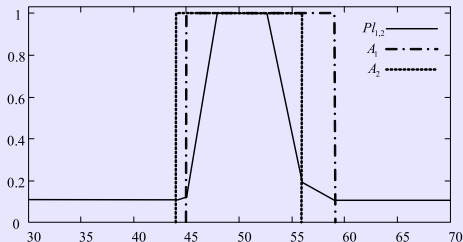
Рассмотрим агрегирование

$$F_{i,j} = (\mathcal{A}_{i,j}, m_{i,j}) = \eta_i F_{A_i} \otimes \eta_j F_{A_j}$$

пар  $(F_1, F_2)$  и  $(F_3, F_6)$  категоричных свидетельств с наибольшим приоритетом с дисконтированием масс и смешанным размытием фокальных элементов.

Визуализировать результат агрегирования можно с помощью **обобщенной контурной функции**

$$Pl_{i,j}(x) = \sum_{A \in \mathcal{A}_{i,j}} \mu_A(x) m_{i,j}(A).$$



Качество агрегирования оценим с помощью анализа изменения следующих характеристик:

- **степень неточности**

$$H(F) = \sum_{A \in \mathcal{A}} m(A) |A|.$$

Пусть  $H_i = H(\eta_i F_{A_i})$ ,  $H_{i,j} = H(F_{i,j})$ ;

- **надежность** вычисляется как нормализованная оценка расстояния

$$d_{i,j} = |E(F_{i,j}) - c_0|$$

от 'среднего' значения прогноза  $E(F_{i,j}) = \sum_{A \in \mathcal{A}_{i,j}} m_{i,j}(A) E(F_A)$  до текущего значения  $c_0$ , где  $E(F_A)$  – центр тяжести нечеткого числа  $A$ ;

- **мера конфликта**

$$Con_k^\Gamma = Con^\Gamma(F_{i,j}, \eta_k F_{A_k}), \quad k = i, j$$

между результатом агрегирования  $F_{i,j}$  и каждым из свидетельств  $\eta_i F_{A_i}$  и  $\eta_j F_{A_j}$ .



	$H_i \downarrow$	$H_j \downarrow$	$H_{i,j} \downarrow$	$\eta_i \uparrow$	$\eta_j \uparrow$	$\eta_{i,j} \uparrow$	$Con^\Gamma \downarrow$	$Con_i^\Gamma \downarrow$	$Con_j^\Gamma \downarrow$
$A_1, A_2$	21.11	23.27	13.77	0.8	0.71	0.73	0.5	0.4	0.34
$A_3, A_6$	27.64	27.51	22.01	0.63	0.61	0.64	0.51	0.65	0.4

Результаты:

- степень неточности после комбинирования уменьшилась;
- надежность возросла для пары  $A_3, A_6$ , а для пары  $A_1, A_2$  – 'по-разному';
- конфликт уменьшился для пары  $A_1, A_2$  и 'по-разному' для пары  $A_3, A_6$ .

# Заклучение

- 1 процедура размытия фокальных элементов, которая моделируется с помощью нечетких чисел, позволяет
  - улучшить робастные свойства вычисления меры конфликта и агрегирования свидетельств;
  - учесть оптимистичность/ радикальность экспертов;
- 2 процедуру размытия фокальных элементов можно совместить с дисконтированием масс, которое отражает степень надежности источников информации – в результате получим преобразование свидетельств;
- 3 параметры преобразований можно выбрать так, чтобы мера конфликта удовлетворяла ряду 'хороших' свойств (монотонность и др.);
- 4 такие преобразования позволяют улучшить результат агрегирования относительно характеристик неточности, надежности и конфликтности.

# СПИСОК ИСТОЧНИКОВ



Лепский, А.Е.: Анализ противоречивости информации в теории функций доверия. Ч. 1. Внешний конфликт. Проблемы управления 5, 3–19 (2021)



Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. 38, 325–339 (1967)



Hodari, D.: Oil prices seen remaining subdued into 2021. The Wall Street Journal, November 28 (2020) <https://www.wsj.com/articles/oil-prices-seen-remaining-subdued-into-2021-11606490833>



Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall PTR (1995)



Lepskiy, A.: Conflict Measure of Belief Functions with Blurred Focal Elements on the Real Line. In: Dencœur T., Lefèvre E., Liu Z., Pichon F. (eds) Belief Functions: Theory and Applications. BELIEF 2021. Lecture Notes in Computer Science, 12915, pp. 197–206. Springer, Cham (2021)



Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press (1976)



Smets, Ph.: Belief functions on real numbers. International Journal of Approximate Reasoning 40(3), 181–223 (2005)



Zhang, L.: Representation, independence and combination of evidence in the Dempster-Shafer theory. In: Yager, R.R. et al. (eds.) Advances in the Dempster-Shafer Theory of Evidence, pp. 51–69. John Wiley & Sons, New York (1994)

**Спасибо за внимание**

alex.lepskiy@gmail.com

alepskiy@hse.ru

<https://www.hse.ru/en/org/persons/10586209>