Об одном методе агрегирования экспертных оценок с учетом надежности и осторожности экспертов

А. Е. Лепский

Национальный исследовательский университет «Высшая школа экономики»

Одной из важных задач экономической информатики является задача агрегирования информации из разных источников. Для этого могут использоваться различные инструментальные методы. Экономическая информация, особенно экспертная, обладает большой степенью неопределенности, неточности, источники информации могут иметь разную надежность. Поэтому математический аппарат, который предполагается использовать для агрегирования такой информации, должен моделировать указанные особенности. В частности, широкий инструментарий для моделирования неопределенности, неточности, надежности источников информации предоставляет теория свидетельств Демпстера — Шейфера (теория функций доверия) [1,2]. Кроме того, в теории функций доверия развит аппарат агрегирования (комбинирования) источников информации, описанных с помощью так называемых тел свидетельств.

В докладе будет показано, как можно методы теории функций доверия использовать для выбора из набора экспертов-источников информации для последующего агрегирования тех, которые наименее конфликтны между собой с учетом надежности этих источников и осторожности экспертов.

Пусть $\mathcal{A}$ – некоторое конечное подмножество непустых множеств из множества $X \subseteq \mathbb{R}$ (множество фокальных элементов). В теории свидетельств рассматривается некоторая неотрицательная функция множеств (функция масс) m, определенная на подмножествах из X, которая удовлетворяет условиям: $m(A) > 0$, если $A \in \mathcal{A}$ и $\sum_{A\in\mathcal{A}} m(A) = 1$. В этом случае пару $F = (\mathcal{A}, m)$ называют телом свидетельств на X.

Телу свидетельств $F = (\mathcal{A}, m)$ взаимно однозначным образом можно поставить в соответствие функции [1,2] доверия $Bel(A) = \sum_{B\subseteq A} m(B)$ и правдоподобия $Pl(A) = \sum_{B: A\cap B\neq\varnothing} m(B)$, которые можно рассматривать как нижнюю и верхнюю оценки вероятности события $\{x \in A\}$.

Тело свидетельств (свидетельство) $F_A = (A,1)$ (т. е. $\mathcal{A} = \{A\}$, $m(A) = 1$) называют категориальным (categorical). В частности, тело свидетельств F_X называют бессодержательным (vacuous). Тогда любое тело свидетельств $F = (\mathcal{A}, m)$ можно представить как выпуклую сумму категориальных свидетельств: $F = \sum_{A\in\mathcal{A}} m(A) F_A$. Например, экспертную информацию {стоимость нефти через полгода будет 78–85 у.е. с достоверностью 0.6, 73–80 у.е. с достоверностью 0.3 или «неизвестно какая»} можно моделировать с помощью тела свидетельств $F = 0.6F_{[78,85]} + 0.3F_{[73,80]} + 0.1F_X$, где (например) $X = [10,110]$ – весь промежуток возможного изменения стоимости нефти. Здесь $\mathcal{A} = \{[78,85],[73,80], X\}$ – множество фокальных элементов. Нетрудно найти, что, например, $Bel([70,82]) = 0.3$, $Pl([70,82]) = 1$.

Будем считать, что все фокальные элементы имеют вид $[a,b)$.

Предположим, что из разных источников (от разных экспертов) мы получили информацию (например, о прогностической стоимости нефти) в виде нескольких тел свидетельств. Мы хотим построить общее тело свидетельств, комбинируя информацию из нескольких источников. Тогда возникают следующие вопросы (задачи): 1) какие тела свидетельств следует выбрать для комбинирования; 2) как комбинировать несколько тел свидетельств в одно; 3) как при решении первых двух задач учесть различную осторожность экспертов и их надежность.

Общей рекомендацией относительно выбора тел свидетельств для комбинирования является выбор малоконфликтных тел свидетельств.

Как оценить конфликт свидетельств? Предположим, что имеется два тела свидетельств $F_1=(\mathcal{A}_1,m_1)$ и $F_2=(\mathcal{A}_2,m_2)$. Необходимо оценить величину конфликта между этими двумя свидетельствами. Для этого будем использовать меру

$$Con^{\Gamma}(F_1,F_2)=\sum_{A\in\mathcal{A}_1,B\in\mathcal{A}_2}\gamma(A,B)m_1(A)m_2(B)\,, \tag{1}$$

где $\Gamma=\left(\gamma(A,B)\right)_{A,B\in\mathcal{A}}$ матрица значений нормированной (т.е. $0\le\gamma(A,B)\le 1$) функции расстояния (меры различия) между промежутками $A,B\in\mathcal{A}$ на $\mathbb{R}$, удовлетворяющую условиям: 1) $\gamma(A,B)=1$, если $A\cap B=\varnothing$; 2) $\gamma(A,A)=0$ $\forall A\subseteq X$. Пусть $|A|$ – мера Лебега множества A на прямой $\mathbb{R}$. Примерами такой функции расстояния являются: а) расстояние Жаккара (Jaccard distance) $\gamma(A,B)=|A\Delta B|/|A\cup B|$; б) Sørensen dissimilarity $\gamma(A,B)=|A\Delta B|/(|A|+|B|)$; 3) $\gamma(A,B)=1$, если $A\cap B=\varnothing$ и $\gamma(A,B)=0$ во всех остальных случаях и др. Заметим, что в случае 3) мы получим в (1) меру $Con_0(F_1,F_2)=\sum_{A\cap B=\varnothing}m_1(A)m_2(B)$, которую называют канонической мерой конфликта (согласованной с правилом комбинирования Демпстера [1]) и чаще всего используют для оценивания конфликта.

Мера конфликта (1) будет согласована с комбинированием тел свидетельств $F_1=(\mathcal{A}_1,m_1)$, $F_2=(\mathcal{A}_2,m_2)$ по правилу $F_S=(\mathcal{A},m_S)=F_1\otimes_S F_2$, где

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

$s(A,B)=1-\gamma(A,B)$, $K=\sum_{A,B}s(A,B)m_1(A)m_2(B)$.

Надежность экспертов можно учитывать с помощью правила дисконтирования Шафера [2]: если число $\eta\in[0,1]$ характеризует степень надежности источника информации ($\eta=1$ – абсолютно надежный источник), то вместо тела свидетельств $F=(\mathcal{A},m)$ будем расссматривать тело свидетельств $F^{(\eta)}=(\mathcal{A},m^{(\eta)})$, где $m^{(\eta)}(A)=\eta\cdot m(A)$ $\forall A\ne X$ и $m^{(\eta)}(X)=1-\eta+\eta\cdot m(X)$.

Для учета различной степени осторожности эксперта (ЛПР) в своих оценках будем вместо фокального элемента $A=[a_1,a_2)$ рассматривать его размытие – нечеткое множество $\tilde{A}$. Возможны различные стратегии размытия в зависимости об информации о степени осторожности ЛПР:

1) если оценки ЛПР, как правило, осторожны, то $\operatorname{supp}\tilde{A}=A$ (внутреннее размытие);
2) если оценки ЛПР, как правило, излишне точны, то $\ker\tilde{A}=A$ (внешнее размытие);
3) если оценки ЛПР нейтральны, то $\mathrm{EI}[\tilde{A}]=\bar{A}$, где EI – ожидаемый интервал нечеткого числа (нейтральное размытие).

В докладе будут проанализированы такие свойства меры конфликта (1) относительно изменений, связанных с дисконтированием масс и размытием фокальных элементов, как: 1) устойчивость к малым изменениям; 2) монотонность относительно изменений; 3) направление изменения.

Кроме того, будет приведен пример применения рассмотренного метода оценивания конфликта с учетом надежности и осторожности экспертов к выбору для агрегирования прогнозов аналитиков относительно стоимости нефти, а также обсуждение качественных результатов агрегирования.

Список литературы

1. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping// The Annals of Mathematical Statistics, 1967, 38(2), 325–339.
2. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976).