УДК 519.816, 004.827

СООТНОШЕНИЕ МЕЖДУ КОНФЛИКТОМ И УМЕНЬШЕНИЕМ НЕЗНАНИЯ ПРИ КОМБИНИРОВАНИИ СВИДЕТЕЛЬСТВ

Лепский А.Е.
Кафедра высшей математики на факультете экономики
Национальный исследовательский университет
«Высшая школа экономики», Москва

Поступила в редакцию 30.07.2013

В работе вводится и исследуется индекс уменьшения незнания после применения правил комбинирования свидетельств в рамках теории Демпстера-Шейфера. Этот индекс анализируются на некоторых специальных множествах (телах) свидетельств. Показано, что достаточным условием уменьшения незнания после применения правил комбинирования является большая корреляция между телами свидетельств. Кроме этого, аксиоматически вводится мера конфликта между свидетельствами. Получен общий вид билинейной меры конфликта. Найдены верхние и нижние оценки зависимости индекса уменьшения незнания от величины меры конфликта после комбинирования свидетельств с помощью правила Демпстера.

The index of decreasing of ignorance after applying of combination rules is introduced and studied in the work within the frame of Dempster-Shafer theory. This index is analysed for some special sets (bodies) of evidence. It is shown that a strong correlation between bodies of evidence is a sufficient condition to decrease of ignorance after applying of combination rules. In addition, measure of conflict between the evidence introduced by axiomatically. A general view of a bilinear measure of conflict found. The upper and lower bounds dependence of the index decreasing of ignorance from the value of measures of the conflict after applying Dempster combination rule found.

Ключевые слова: теория Демпстера-Шейфера, правила комбинирования, индекс неточности, мера конфликта.
Keywords: Dempster-Shafer theory, combination rules, imprecision index, conflict measure.

1. Введение

В теории Демпстера-Шейфера [1, 2] (теории свидетельств, теории функций доверия) рассматриваются правила комбинирования свидетельств для агрегирования экспертной информации, полученной из разных источников. Исторически первым таким правилом было правило Демпстера [1]. Это правило было подвергнуто, как многочисленному анализу, так и критике [3]-[8]. В результате этого были предложены новые правила комбинирования. Эти правила применимы в тех или иных ситуациях, обладают разными свойствами. В одних ситуациях комбинирования то или иное правило дает правдоподобный результат, а в других ситуациях – неправдоподобный. Более того, правдоподобность комбинирования часто зависит только от контекста, а не от конкретных параметров комбинируемых свидетельств. Но в первую очередь это связано со следующими причинами: 1) возможной конфликтностью свидетельств; 2) возможностью большого недостатка информации, содержащейся в свидетельствах (незнание свидетельств); 3) различной интерпретируемостью свидетельств. Ниже мы не будем анализировать третью причину, но остановимся на исследовании первых двух причин.

Эффективность применения правил комбинирования может быть оценена с помощью оценки изменения количества незнания после применения правил комбинирования. В этой работе количество незнания будет оценено с помощью, так называемых индексов неточности, определенных на монотонных мерах (например, функциях доверия). Такие индексы исследовались во многих работах в рамках теории неточных вероятностей. В этой статье будет использовано аксиоматическое определение индекса неточности, предложенное в работах [9, 10]. Предположим, что мы используем некоторое правило комбинирования R для комбинирования двух свидетельств. Как результат мы получим новое свидетельство. Тогда можно поставить вопрос о количестве уменьшения незнания после применения правила комбинирования R. Это количество может быть оценено с помощью индекса неточности как величина изменения распределения элементов тела свидетельств. Существуют также различные подходы к определению меры конфликта свидетельств (или соответствующих функций доверия). Например, хорошо известен метрический подход к определению меры конфликта как расстояния между двумя базовыми вероятностными назначениями [11, 12]. В этой работе мера конфликта будет введена аксиоматически как функционал, определенный на декартовом произведении множеств мер доверия и обладающий свойством антимонотонности относительно степени пересечения множеств. Вводимые меры конфликта и мера изменения незнания будут исследованы на специальных типах тел свидетельств. Для таких тел, в частности будут найдены условия, гарантирующие уменьшение незнания после применения тех или иных правил комбинирования.

Некоторые результаты этой работы были анонсированы в [13].

1. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ДЕМПСТЕРА-ШЕЙФЕРА

1.1 Основные определения и обозначения

Основными понятиями теории Демпстера-Шейфера (теории свидетельств) являются понятия функций доверия и правдоподобия. Пусть X – конечное универсальное множество и 2^X множество всех подмножеств из X. Рассмотрим *меру доверия* (или *функцию доверия*)[14] $g : 2^X \to [0, 1]$. В теории свидетельств [2] значение $g(A)$, $A \in 2^X$, интерпретируется как степень доверия к тому, что истинная альтернатива из X принадлежит множеству A [14]. Функция доверия g определяется с помощью функции множеств $m_g(A)$, которая называется *базовым вероятностным назначением* (бвн). Эта функция должна удовлетворять следующим условиям [2]:

$$m_g : 2^X \to [0,1], \;\; m_g(\emptyset) = 0, \;\; \sum_{A \subseteq X} m_g(A) = 1. \tag{1}$$

Тогда $g(A) = \sum_{B:\, B \subseteq A} m_g(B)$. Обозначим множество всех функций доверия на 2^X через $Bel(X)$. Вместе с функцией доверия g в теории свидетельств рассматривается также и двойственная ей функция $\bar{g}$, которая называется *функцией правдоподобия* и вычисляется как $\bar{g}(A) = 1 - g(\bar{A})$, $A \in 2^X$. Обратно, базовое вероятностное назначение m_g может быть вычислено по функции доверия g с помощью, так называемого *преобразования Мёбиуса*: $m_g(B) = \sum_{A:A \subseteq B} (-1)^{|B \setminus A|} g(A)$. Функции доверия и правдоподобия могут рассматриваться как нижняя и верхняя оценки вероятности события. Действительно, для любой меры доверия g может быть найдена такая вероятность p, что $g(A) \leq p(A) \leq \bar{g}(A)$ для всех $A \in 2^X$. Тогда пара $(g(A), \bar{g}(A))$ характеризует неопределенность вероятности события A. Функция доверия имеет следующую статистическую интерпретацию. Пусть N экспертов высказали свои мнения (свидетельства) о значениях некоторой переменной $x \in X$. Причем c_i, $i = 1, ..., k$, экспертов из N предположили, что $x \in A_i$, $i = 1, ..., k$, где $A_i \in 2^X$. Имеем $\sum_{i=1}^{k} c_i = N$. Пусть $m(A_i) = c_i / N$ – частота события $\{x \in A_i\}$, $i = 1, ..., k$. Подмножество $A \in 2^X$ называется *фокальным элементом*, если $m(A) > 0$. Пусть $\mathcal{A}$ – множество всех фокальных элементов. Тогда функция множеств $m(A)$, если $A \in \mathcal{A}$ и $m(A) = 0$ в противном случае удовлетворяет условию (1). Пара $F = (\mathcal{A}, m)$ называется *телом свидетельств*. Будем обозначать через $\mathcal{A}(g)$ и $F(g)$ множества всех фокальных элементов и тело свидетельств соответственно, связанных с функцией доверия g.

Замечание 1. Без усложнения дальнейшего изложения можно рассматривать также свидетельства, имеющие конечное множество фокальных элементов, определенных на бесконечном базовом множестве X (например, $X = \mathbb{R}$ – числовая прямая).

1.2 Правила комбинирования свидетельств

Предположим, что есть два тела свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$, связанных с функциями доверия $g_1, g_2 \in Bel(X)$. Например, это могут быть свидетельства, полученные от двух источников информации. Тогда может быть поставлена задача комбинирования этих двух свидетельств в одно свидетельство. В рамках теории Демпстера-Шейфера рассматриваются различные правила комбинирования. В общем случае правило комбинирования – это некоторый оператор $R : Bel(X) \times Bel(X) \to Bel(X)$. Детальный и критичный обзор правил комбинирования можно найти в отчете [15]. Ниже рассмотрим только несколько основных правил комбинирования.

а) Правило Демпстера. Это правило было введено в работе [1] и обобщено в [2] для комбинирования произвольных независимых свидетельств. Оно определяется следующим образом

$$m_D(A) = \frac{1}{1-K} \sum_{A_1 \cap A_2 = A} m_{g_1}(A_1) m_{g_2}(A_2), \;\; A \neq \emptyset; \;\; m_D(\emptyset) = 0;$$

$$K = K(g_1, g_2) = \sum_{A_1 \cap A_2 = \emptyset} m_{g_1}(A_1) m_{g_2}(A_2). \tag{2}$$

Величина $K(g_1, g_2)$ характеризует количество конфликта в двух источниках информации, которые задаются с помощью тел свидетельств $F(g_1)$ и $F(g_2)$. Если $K(g_1, g_2) = 1$, то это означает, что источники информации абсолютно конфликтны и правило Демпстера не может быть применено. Правило Демпстера было подвергнуто многочисленной критике [3]-[8], в результате которой были предложены новые правила комбинирования.

b) Правило дисконтирования. Это правило было введено Шафером [2]. Основная идея состояла в использовании некоторого коэффициента $\alpha \in [0,1]$ для дисконтирования базовых вероятностей: $m^{\alpha}(A) = (1-\alpha)m(A)$, $A \neq X$, $m^{\alpha}(X) = \alpha + (1-\alpha)m(X)$. Этот коэффициент характеризует степень надежности информации: если $\alpha = 0$, то источник информации считается абсолютно надежным. Если же $\alpha = 1$, то источник информации абсолютно ненадежен. После дисконтирования базовых вероятностных назначений применяется правило Демпстера (2). Если $\alpha \in (0,1)$, то правило дисконтирования (2) может применяться для любых тел событий (в этом случае значение $K \neq 0$).

c) Правило комбинирования Ягера. Это правило было введено в работе [4] и определяется следующим образом

$$q(A) = \sum_{A_1 \cap A_2 = A} m_{g_1}(A_1) m_{g_2}(A_2), \quad A \in 2^X, \tag{3}$$

$$m_Y(A) = q(A), \ A \neq \emptyset, X; \ \ m_Y(\emptyset) = q(\emptyset) = K, \ \ m_Y(X) = m_Y(\emptyset) + q(X),$$

где значение $K = K(g_1, g_2)$ определяется так же, как и в (2). Значение $q(X) = m_{g_1}(X) m_{g_2}(X)$ характеризует количество незнания в двух телах свидетельств $F(g_1)$ и $F(g_2)$. Поэтому правило Ягера использует информацию о конфликте (значение $q(\emptyset) = K$) и незнании (значение $q(X)$) только при вычислении базовых вероятностных назначений универсального множества X. Этот подчеркивает «осторожность» этого правила.

d) Модифицированное правило комбинирования Инагаки [16]. Это правило определяется с помощью функции $q(A)$, которая использовалась Ягером [4] в (3) и неотрицательного параметра k:

$$m_I(A) = q(A)(1 + kq(\emptyset)), \ A \neq X; \ \ m_I(X) = q(X)(1 + kq(\emptyset)) + q(\emptyset)(1 + kq(\emptyset) - k),$$

где $0 \leq k \leq 1/(1 - q(\emptyset) - q(X))$. Если $k = 0$, то мы получим правило Ягера. Если $k = 1/(1 - q(\emptyset))$, то получим правило Демпстера. Поэтому правило Инагаки использует информацию о конфликте и незнании при вычислении базовых вероятностных назначений всех множеств с коэффициентом $(1 + kq(\emptyset))$, который регулирует соотношение между конфликтом и незнанием.

e) Правило комбинирования Цанга. Это правило было введено в работе [17] и задается следующим образом

$$m_Z(A) = \sum_{A_1 \cap A_2 = A} r(A_1, A_2) m_{g_1}(A_1) m_{g_2}(A_2), \quad A \in 2^X,$$

где $r(A_1, A_2)$ – мера пересечения множеств A_1 и A_2. Например, $r(A_1, A_2) = c\frac{|A_1 \cap A_2|}{\min\{|A_1|, |A_2|\}}$ – подобность Симпсона [18] или $r(A_1, A_2) = c\frac{|A_1 \cap A_2|}{|A_1 \cup A|}$ – подобность Жаккарда [18], где $c > 0$ – нормирующий множитель.

f) Дизъюнктивное правило консенсуса Дюбуа-Прада [19]:

$$m_{DP}(A) = \sum_{A_1 \cup A_2 = A} m_{g_1}(A_1) m_{g_2}(A_2), \quad A \in 2^X.$$

Существуют и другие правила комбинирования. Приведенные примеры правил комбинирования показывают, что при комбинировании свидетельств важно учитывать информацию о конфликте и незнании этих свидетельств. Ниже будут определены функционалы, с помощью которых может быть оценено количество конфликта и незнания в конкретной ситуации комбинирования свидетельств.

2. ИЗМЕНЕНИЕ НЕЗНАНИЯ

Эффективность применения правил комбинирования может быть оценено количеством уменьшения незнания после применения того или иного правила комбинирования. Для вычисления количества незнания будем использовать понятие индекса неточности. В общем случае индекс неточности f определяется на множестве $Bel(X)$ и характеризует степень отклонения (меру неопределенности) функции доверия g от вероятностной меры. Ниже будет показано, что значение меры неопределенности можно рассматривать как показатель информационного незнания, содержащейся в мере g. То или иное правило комбинирования есть смысл применять только в том случае, если после его применения уровень незнания уменьшится. Степень уменьшения незнания может быть оценена с помощью сравнения $f(g)$ с $f(g_1)$ и $f(g_2)$, где g – функция доверия после комбинирования свидетельств $F(g_1)$ и $F(g_2)$.

2.1 Индекс неточности

Измерение неопределенности играет важную роль в различных «неопределенностных» теориях: теории вероятностей, теории возможностей, теории свидетельств, теории информации, теории нечетких множеств и т.д. Существует несколько подходов к определению меры неопределенности в теории свидетельств. Ниже будем придерживаться подхода, который был введен в работах [9, 10]. Предположим, что нам только известно, что истинная альтернатива принадлежит непустому множеству $B \subseteq X$. Эта ситуация может быть описана с помощью неаддитивной меры (так называемой примитивной меры доверия)

$$\eta_{\langle B \rangle}(A) = \begin{cases} 1, & B \subseteq A, \\ 0, & B \not\subseteq A, \end{cases}, \quad A \subseteq X, \quad B \neq \emptyset,$$

которая дает нижнюю вероятность события A. Степень неопределенности такой меры описывается с помощью известной *меры Хартли* $H(\eta_{\langle B \rangle}) = \log_2 |B|$, которая характеризует степень неточности информации о принадлежности «истинной» альтернативы множеству $B \subseteq X$.

Обобщением этой ситуации является следующая конструкция. Пусть $g \in Bel(X)$. Рассмотрим пару $(g, \bar{g})$, $g(A) \leq \bar{g}(A)$ для всех $A \in 2^X$, $g(\emptyset) = \bar{g}(\emptyset) = 0$. Предположим, что существует такая вероятностная мера P на 2^X, что $g(A) \leq P(A) \leq \bar{g}(A)$ для всех $A \in 2^X$. Другими словами, функции множеств $g, \bar{g}$ дают нам нижнюю и верхнюю границы вероятности, и для любого события $A \in 2^X$ мы имеем только интервал $[\underline{g}(A), \bar{g}(A)]$ возможных значений «истинной» вероятности $P(A)$.

Существует обобщение меры Хартли. Пусть g функция доверия, которая может быть представленная в виде $g = \sum_{B \in 2^X} m(B)\eta_{\langle B\rangle}$, где $m(\emptyset) = 0$, $m(B) \geq 0$ для всех $B \in 2^X$ и $\sum_{B \in 2^X} m(B) = 1$. Тогда *обобщенная мера Хартли* определяется как

$$GH\,(g) = \sum_{B \in 2^X \backslash \{\emptyset\}} m(B) \log_2 |B|\,. \tag{4}$$

Замечание 2. Обобщенную меру Хартли можно рассматривать и в том случае, когда функция доверия g задана на конечном множестве фокальных элементов бесконечного множества X. Если X – множество конечной меры некоторого пространства с мерой $(E, \mathcal{F}, \mu)$ и $A \subseteq X$, $A \in \mathcal{F}$, то в качестве $|A|$ можно использовать $|A| = 1 + \mu(A)$.

Определение 1 [9]. Функционал $f : Bel(X) \to [0,1]$ называется *индексом неточности*, если он удовлетворяет следующим условиям: 1) $f(g) = 0$, если g – вероятностная мера; 2) $f(g_1) \geq f(g_2)$ для всех таких $g_1, g_2 \in Bel(X)$, что $g_1 \leq g_2$ (т.е. $g_1(A) \leq g_2(A)$ для всех $A \in 2^X$); 3) $f\left(\eta_{\langle X\rangle}\right) = 1$. Индекс неточности f на $Bel(X)$ называется *линейным*, если для любой линейной комбинации $\sum_{j=1}^{k} \alpha_j g_j \in Bel(X)$, $\alpha_i \in \mathbb{R}$,$g_j \in Bel(X)$, $j = 1, ..., k$, имеем $f\left(\sum_{j=1}^{k} \alpha_j g_j\right) = \sum_{j=1}^{k} \alpha_j f\,(g_j)$.

Замечание 3. Будем писать $g_1 < g_2$ для $g_1, g_2 \in Bel(X)$, если $g_1 \leq g_2$ и $g_1 \neq g_2$.

Заметим, что функция доверия g может быть представлена как линейная комбинация примитивных функций доверия $\eta_{\langle B\rangle}$:

$$g = \sum_{B \in 2^X \backslash \{\emptyset\}} m_g(B)\eta_{\langle B\rangle}. \tag{5}$$

Из формулы (5) видно, что множество $\{\eta_{\langle B\rangle}\}$, $B \in 2^X \backslash \{\emptyset\}$, является базисом в множестве $Bel(X)$ в том смысле, что любая функция доверия $g \in Bel(X)$ представляется единственным образом как выпуклая комбинация примитивных мер $\{\eta_{\langle B\rangle}\}$, $B \in 2^X \backslash \{\emptyset\}$. С другой стороны, любой линейный функционал f на $Bel(X)$ определяется единственным образом своими значениями на указанном базисе в $Bel(X)$. Это позволяет определить f с помощью функции множеств в соответствии со следующим правилом $\mu_f(B) = f\left(\eta_{\langle B\rangle}\right)$, $B \in 2^X \backslash \{\emptyset\}$. Положим по определению что $\mu_f(\emptyset) = 0$ для любого индекса неточности f. В работе [10] были найдены различные представления индекса неточности. В этой статье будем использовать следующее представление.

Предложение 1. Функционал $f : Bel(X) \to [0,1]$ является линейным индексом неточности на $Bel(X)$ тогда и только тогда, когда

$$f(g) = \sum_{B \in 2^X \backslash \{\emptyset\}} m_g(B)\mu_f(B), \tag{6}$$

где монотонная неотрицательная функция множеств μ_f удовлетворяет условиям: 1) $\mu_f(\{x\}) = 0$ для любых $x \in X$; 2) $\mu_f(X) = f\left(\eta_{\langle X\rangle}\right) = 1$; 3) $\sum_{B: A \subseteq B} (-1)^{|B \backslash A|} \mu_f(B) \leq 0$ для всех $A \neq \emptyset, X$.

Обобщенная нормализованная мера Хартли $GH_0 = GH/\log_2 |X|$ (см. (4)) является примером линейного индекса неточности. Формулы (4) и (6) показывают, что линейный индекс неточности f определяет некоторое распределение на теле свидетельств. Это распределение имеет плотность μ_f. Большие значения плотности соответствуют фокальным элементам, имеющим большую мощность (см. свойство 3) в предложении 1). Наличие больших по мощности и весу свидетельств характеризует большую степень незнания. Поэтому значение линейного индекса неточности $f(g)$ оценивает степень такого незнания.

Замечание 4. Нетрудно показать, что множество всех линейных индексов неточности является выпуклым.

2.2 Индекс уменьшения незнания

Предположим, что мы имеем два тела свидетельств $F(g_1) = (\mathcal{A}^{(1)}, m^{(1)})$ и $F(g_2) = (\mathcal{A}^{(2)}, m^{(2)})$, которые определены на множестве X и отвечают двум функциям доверия g_1 и g_2 соответственно. Пусть $f : Bel(X) \to [0,1]$ некоторый линейный индекс неточности, с помощью которого оценивается степень незнания, содержащейся в информации, задаваемой функцией g. Предположим, что мы используем некоторое правило комбинирования R для агрегирования свидетельств $F(g_1)$ и $F(g_2)$. В результате мы получим новую функцию доверия $g = R(g_1, g_2)$. Тогда может быть поставлен вопрос о величине уменьшения незнания после применения правила комбинирования R. Величина такого уменьшения может быть оценена с помощью сравнения $f(g)$ с $f(g_1)$ и $f(g_2)$. Например, может быть введен следующий индекс уменьшения незнания

$$I_R(g_i|g_j) = f(g_i) - f(R(g_i, g_j)), \quad i, j \in \{1, 2\},$$

$$I_R(g_1, g_2) = \min\left\{I_R(g_1|g_2), I_R(g_2|g_1)\right\}.$$

Уменьшение незнания соответствует случаю положительности индекса $I_R(g_1, g_2)$.

Рассмотрим изменение незнания при комбинировании тел событий определенных видов. Важным типом пар тел свидетельств при комбинировании является пара следующего вида, которую назовем парой *согласованных тел свидетельств.*

Предположим, что множества фокальных элементов $\mathcal{A}(g_1)$ и $\mathcal{A}(g_2)$, связанные с функциями доверия $g_1, g_2 \in Bel(X)$, удовлетворяют условиям:

1) $A' \cap A'' = \emptyset$, $B' \cap B'' = \emptyset$ для всех $A', A'' \in \mathcal{A}(g_1)$, $B', B'' \in \mathcal{A}(g_2)$;

2) для любого множества $A \in \mathcal{A}(g_1)$ найдется такое единственное множество $B \in \mathcal{A}(g_2)$, что $A \cap B \neq \emptyset$;

3) для любого множества $B \in \mathcal{A}(g_2)$ найдется такое единственное множество $A \in \mathcal{A}(g_1)$, что $A \cap B \neq \emptyset$.

Согласованные свидетельства появляются, например, в тех случаях, когда два эксперта дают информацию в «заранее очерченных рамках» (например, оптимистичный и пессимистичный сценарии развития какой-либо ситуации).

Таким образом, для пары согласованных тел свидетельств существует взаимно однозначное соответствие φ между элементами множеств $\mathcal{A}(g_1)$ и $\mathcal{A}(g_2)$. Если кроме условий 1)-3), множества фокальных элементов удовлетворяют также условию

4) $A \subseteq \varphi(A)$ для всех $A \in \mathcal{A}(g_1)$, то такую ситуацию будем называть *уточняющим телом свидетельств* (тело свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ уточняет тело свидетельств $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$).

Поставим вопрос об уменьшении незнания после применения правила комбинирования к паре согласованных тел свидетельств. Оказывается, что ответ на этот вопрос зависит от того, какое правило комбинирования используется. Сформулируем результат об уменьшении незнания после применения правила комбинирования Демпстера к паре согласованных тел свидетельств.

Предложение 2. Пусть $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ два тела свидетельств, удовлетворяющие условиям 1)-3). Тогда уменьшения незнания $I_D(g_1, g_2)$ будет положительным для правила Демпстера, если выполняется следующее условие:

$$\sum_{A\in\mathcal{A}(g_1)} m_{g_1}(A)m_{g_2}(\varphi(A)) >$$

$$\max_{A\in\mathcal{A}(g_1)} \frac{\mu_f(A\cap\varphi(A))}{\mu_f(A)\mu_f(\varphi(A))} \max\left\{m_{g_1}(A)\mu_f(A), m_{g_2}(\varphi(A))\mu_f(\varphi(A))\right\}. \tag{7}$$

Условие (7) несколько упрощается в случае комбинирования уточняющих тел свидетельств с помощью правила Демпстера.

Следствие 1. Пусть два тела свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ удовлетворяют условиям 1)-4). Тогда индекс уменьшения незнания $I_D(g_1, g_2)$ будет положительным для правила Демпстера, если справедливо следующее условие:

$$\sum_{A\in\mathcal{A}(g_1)} m_{g_1}(A)m_{g_2}(\varphi(A)) > \max_{A\in\mathcal{A}(g_1)} \max\left\{m_{g_1}(A)\frac{\mu_f(A)}{\mu_f(\varphi(A))}, m_{g_2}(\varphi(A))\right\}. \tag{8}$$

Замечание 5. Выражение слева в (7) (или (8)) представляет собой аналог «скалярного произведения» двух тел векторов-свидетельств (или корреляцию между двумя телами свидетельств). Поэтому это выражение имеет тем большее значение, чем «более коллинеарны» (или согласованы) вектора-свидетельства. Таким образом, неравенство (7) (или (8)) означает, что большая корреляция свидетельств является достаточным условием для уменьшения незнания после применения правила комбинирования Демпстера. Кроме того, для тел согласованных (и уточняющих) свидетельств величина конфликта $K(g_1, g_2)$ будет вычисляться по формуле

$$K(g_1, g_2) = 1 - \sum_{A\in\mathcal{A}(g_1)} m_{g_1}(A)m_{g_2}(\varphi(A)).$$

Поэтому условиям (7) или (8) можно интерпретировать так: для уменьшения незнания после применения правила комбинирования Демпстера достаточно, чтобы конфликтность свидетельств $K(g_1, g_2)$ была не очень большой. В тоже время, корреляция между векторами-свидетельствами должна быть достаточно большой. Даже при комбинировании одинаковых тел свидетельств условие (7) (и (8)) не выполняется – комбинирование с помощью правила Демпстера одинаковых свидетельств не уменьшает количество незнания. Более того, это условие не выполняются, если согласованные фокальные элементы тел свидетельств – одинаковы, а базовые вероятностные назначения – различны. Условие (7) в этом случае примет вид: $\sum_{A\in\mathcal{A}(g_1)} m_{g_1}(A)m_{g_2}(A) > \max\limits_{A\in\mathcal{A}(g_1)} \max\left\{m_{g_1}(A), m_{g_2}(A)\right\}$. Другими словами, для уменьшения количества незнания согласованные фокальные элементы комбинируемых свидетельств должны быть различны.

Пример 1. Эксперты двух агентств делают прогноз о стоимости акций некоторой компании, предполагая два сценария экономического развития. В первом агентстве 6 экспертов высказались за стоимость акций в промежутке $[5, 7)$ у.е., предполагая стагнацию экономики, а 4 эксперта – $[9,10)$ в предположении экономического роста. Во втором агентстве в первом случае 5 экспертов дали прогноз $[5,8)$, а 3 эксперта – $[8,10)$. Оценим конфликт и уменьшение незнания при комбинировании с помощью правила Демпстера этих свидетельств.

Решение. В данном случае имеем два уточняющих тела свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$, где $\mathcal{A}(g_1) = \{A_1, A_2\}$, $\mathcal{A}(g_2) = \{B_1, B_2\}$, причем $A_1 = [5,6)$, $A_2 = [9,11)$, $B_1 = [5,7)$, $B_2 = [8,11)$, $m_{g_1}(A_1) = \frac{6}{10}$, $m_{g_1}(A_2) = \frac{4}{10}$, $m_{g_2}(B_1) = \frac{5}{8}$, $m_{g_2}(B_2) = \frac{3}{8}$. Пусть $X = [5,11)$ и для $A = [a,b) \subseteq X$ положим $|A| = 1 + b - a$ (см. Замечание 2). Величина конфликта тел свидетельств $F(g_1)$ и $F(g_2)$ будет равна $K = K(g_1, g_2) = \frac{6}{10} \cdot \frac{3}{8} + \frac{4}{10} \cdot \frac{3}{8} = \frac{19}{40}$. После применения правила Демпстера получим комбинированное свидетельство $F(g) = (\mathcal{A}(g), m_g)$ с $\mathcal{A}(g) = \mathcal{A}(g_1)$ и $m_g(A_1) = \frac{\frac{6}{10} \cdot \frac{5}{8}}{1-K} = \frac{5}{7}$, $m_g(A_2) = \frac{2}{7}$. Найдем значения индексов неточности $f(g)$ и $f(g_i)$, $i = 1,2$, с помощью обобщенной меры Хартли. Имеем

$$f(g_1) = \tfrac{6}{10} \ln 2 + \tfrac{4}{10} \ln 3, \quad f(g_2) = \tfrac{5}{8} \ln 3 + \tfrac{3}{8} \ln 4, \quad f(g) = \tfrac{5}{7} \ln 3 + \tfrac{2}{7} \ln 2.$$

Тогда $f(g_1) - f(g) = -\frac{4}{35} \ln 2 + \frac{4}{35} \ln 3 > 0$, $f(g_2) - f(g) = -\frac{5}{56} \ln 3 + \frac{26}{56} \ln 2 > 0$. Таким образом, индекс уменьшения незнания будет равен $I_D(g_1, g_2) = \min\{f(g_i) - f(g)\} = \frac{4}{35} \ln \frac{3}{2}$. Другими словами, количество незнания после комбинирования свидетельств по правилу Демпстера уменьшится на величину $\frac{4}{35} \ln \frac{3}{2}$, т.е. примерно на 4,6

Рассмотрим теперь вопрос об изменении индекса незнания при использовании правила комбинирования Ягера.

Предложение 3. Пусть $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ два тела свидетельств, удовлетворяющие условиям 1)-3). Тогда индекс уменьшения незнания $I_Y(g_1, g_2)$ будет положительным для правила Ягера тогда и только тогда, когда

$$\sum_{A \in \mathcal{A}(g_1)} m_{g_1}(A) m_{g_2}(\varphi(A)) \left(1 - \mu_f(A \cap \varphi(A))\right) >$$

$$\max \left\{ \sum_{A \in \mathcal{A}(g_1)} m_{g_1}(A) \left(1 - \mu_f(A)\right), \sum_{A \in \mathcal{A}(g_1)} m_{g_2}(\varphi(A)) \left(1 - \mu_f(\varphi(A))\right) \right\}.$$

Однако индекс незнания не может быть уменьшен с помощью правила Ягера при комбинировании уточняющих тел свидетельств.

Следствие 2. Пусть два тела свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ удовлетворяют условиям 1)-4). Тогда индекс уменьшения незнания $I_Y(g_1|g_2) = f(g_1) - f(Y(g_1, g_2))$ будет неположительным для правила Ягера.

3. МЕРА КОНФЛИКТА

Пусть $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ – два тела свидетельств на X, связанные с функциями доверия g_1 и g_2 соответственно. Величина $K(g_1, g_2)$, вычисляемая по формуле (2) и численно равная сумме произведений базовых вероятностных назначений непересекающихся фокальных элементов двух тел свидетельств, характеризует количество конфликта в двух источниках информации, которые задаются с помощью тел свидетельств $F(g_1)$ и $F(g_2)$. Если $K(g_1, g_2) = 1$, то это означает, что в двух источниках информации нет свидетельств, содержащих общие элементы. В этом случае можно считать, что источники информации абсолютно конфликтны. Правило Демпстера в такой ситуации неприменимо. Величину $K(g_1, g_2)$ будем называть *канонической мерой конфликта*, связанного с функциями доверия g_1 и g_2. Однако конфликтность двух источников информации будет большой и в том случае, если эти источники представлены свидетельствами, у которых есть общие элементы, но их количество мало по сравнению с мощностью сравниваемых фокальных элементов. Поэтому аксиоматически введем понятие меры конфликта тел свидетельств $F(g_1)$ и $F(g_2)$, отражающих указанное наблюдение. В общем случае такая мера будет зависеть от степени пересечения двух множеств.

Пусть $r : 2^X \times 2^X \to [0, 1]$ – некоторый коэффициент пересечения множеств, удовлетворяющий условиям:

1) $r(A, B) = r(B, A)$;

2) $r(A, B) = 0$, если $A \cap B = \emptyset$;

3) $r(A, A) = 1$, $A \neq \emptyset$.

Примером коэффициента пересечения множеств является так называемая *подобность перекрытия Симпсона* [18] $r(A, B) = \frac{|A \cap B|}{\min\{|A|, |B|\}}$, если $A \neq \emptyset$ и $B \neq \emptyset$.

Определение 2. Функционал $c_r : Bel(X) \times Bel(X) \to [0, 1]$ называется *мерой конфликта* относительно коэффициента пересечения множеств r, если он удовлетворяет следующим условиям:

1) $c_r(g_1, g_2) = c_r(g_2, g_1)$ для всех $g_1, g_2 \in Bel(X)$;

2) $c_r(g', g) \geq c_r(g'', g)$, если $F(g') = F \cup (A', m)$, $F(g'') = F \cup (A'', m)$ и $r(A', B) \leq r(A'', B)$ для всех $B \in \mathcal{A}(g)$;

3) $c_r(g_1, g_2) = 1$, если $A \cap B = \emptyset$ для всех $A \in \mathcal{A}(g_1)$, $B \in \mathcal{A}(g_2)$.

Первая аксиома это условие симметричности меры относительно двух источников свидетельств. Вторая аксиома – условие антимонотонности меры конфликта относительно степени пересечения двух свидетельств: чем меньше степень пересечения, тем больше конфликт. Третья аксиома – условие нормировки: мера конфликта равна единице в случае наиболее конфликтной ситуации, когда два тела свидетельств не имеют пересекающихся фокальных элементов. Алгебраически наиболее простой мерой конфликта двух тел свидетельств (или соответствующих функций доверия) является мера, линейная по каждому аргументу. Введем соответствующее понятие.

Мера конфликта c_r на $Bel(X) \times Bel(X)$ называется *билинейной*, если $c_r(\alpha g_1 + \beta g_2, g) = \alpha c_r(g_1, g) + \beta c_r(g_2, g)$ для всех $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$, $g, g_1, g_2 \in Bel(X)$.

Предложение 4. Функционал c_r билинейная мера конфликта на $Bel(X) \times Bel(X)$ тогда и только тогда, когда

$$c_r(g_1, g_2) = \sum_{A \in \mathcal{A}(g_1), B \in \mathcal{A}(g_2)} \gamma(A, B) m_{g_1}(A) m_{g_2}(B) =$$

$$K(g_1, g_2) + \sum_{A \cap B \neq \emptyset} \gamma(A, B) m_{g_1}(A) m_{g_2}(B), \tag{9}$$

где $\gamma(A, B) \in [0, 1]$ удовлетворяют условиям:

а) $\gamma(A, B) = \gamma(B, A)$;

б) $\gamma(A', B) \geq \gamma(A'', B)$, если $r(A', B) \leq r(A'', B)$;

в) $\gamma(A, B) = 1$, если $A \cap B = \emptyset$.

Из формулы (9) следует, что каноническая мера конфликта $K(g_1, g_2)$ является наименьшей из всех билинейных мер конфликта: $c_r(g_1, g_2) \geq K(g_1, g_2)$. Например, коэффициенты $\gamma(A,B) = c_r\left(\eta_{\langle A\rangle}, \eta_{\langle B\rangle}\right) = \psi(r(A,B))$, $A,B \neq \emptyset$, будут удовлетворять условиям а) – в) Предложения 4, если ψ невозрастающая функция, $\psi(1) = 0$, $\psi(0) = 1$ и $r(A,B) = \frac{|A\cap B|}{\min\{|A|,|B|\}}$. В частности, если $r(A,B) = \begin{cases} 1, & A\cap B \neq \emptyset, \\ 0, & A\cap B = \emptyset \end{cases}$ – примитивная мера пересечения, то $c_r(g_1,g_2) = K(g_1,g_2)$.

Мера конфликта может быть использована для предварительного оценивания количеств конфликта двух свидетельств.

Пример 2. Пусть $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ являются двумя согласованными телами свидетельств (см. 3.2). Тогда

$$c_r(g_1,g_2) = 1 - \sum_{A\in\mathcal{A}(g_1)} (1-\gamma(A,\varphi(A)))m_{g_1}(A)m_{g_2}(\varphi(A)).$$

В частности, если

$$\gamma(A,B) = 1 - r(A,B) = 1 - |A\cap B|/\min\{|A|,|B|\},$$

то

$$c_r(g_1,g_2) = 1 - \sum_{A\in\mathcal{A}(g_1)} \frac{|A\cap\varphi(A)|}{\min\{|A|,|\varphi(A)|\}} m_{g_1}(A)m_{g_2}(\varphi(A)).$$

Если же мы имеем два уточняющих тела свидетельств, то последнее выражение упрощается следующим образом

$$c_r(g_1,g_2) = K = 1 - \sum_{A\in\mathcal{A}(g_1)} m_{g_1}(A)m_{g_2}(\varphi(A)).$$

4. ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ МЕЖДУ МЕРОЙ КОНФЛИКТА И ИНДЕКСОМ УМЕНЬШЕНИЯ НЕЗНАНИЯ

Исследуем зависимость между мерой конфликта и индексом уменьшения незнания в случае комбинирования разных видов свидетельств и применения различных правил комбинирования. Возможно как аналитическое исследование указанной зависимости, так и вероятностное в предположении, что комбинированные свидетельства определенного вида выбираются в соответствии с некоторым вероятностным законом. В последнем случае оценки вероятностных распределений и характеристик можно получить методом статистических испытаний.

4.1 Статистическое исследование

Проведем исследование зависимости между мерой конфликта и индексом уменьшения незнания в случае комбинирования по правилу Демпстера двух тел свидетельств $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$, причем $|\mathcal{A}(g_1)| = |\mathcal{A}(g_2)| = 2$. Рассмотрим три вида комбинируемых тел свидетельств: 1) общий вид; 2) согласованные свидетельства; 3) уточняющие свидетельства. В качестве меры конфликта будем использовать величину $K = K(g_1,g_2)$ – каноническую меру конфликта, вычисляемую по формуле (2). Для статистического исследования осуществим равномерную генерацию фокальных элементов свидетельств определенного вида и базовых вероятностных назначений. Приведем некоторые результаты статистического исследования.

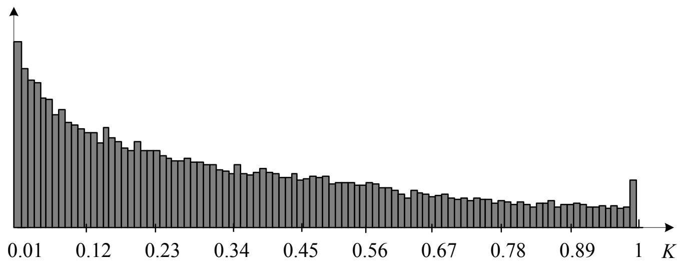

Рис. 1: Гистограмма меры конфликта в общем случае

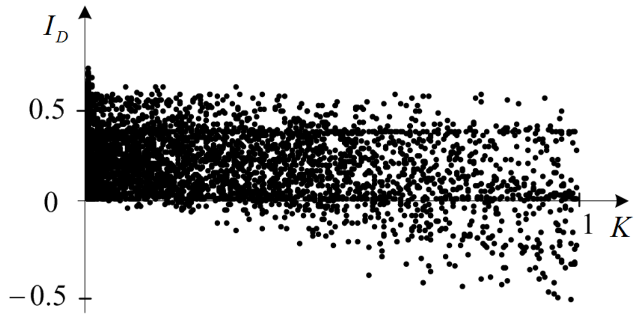

Рис. 2: Распределение точек (K, I_D) при комбинировании двух свидетельств в общем случае

1. Общий случай тел свидетельств.

В общем случае количество сильно конфликтных свидетельств больше числа слабо конфликтных свидетельств (см. Рис.1).

Распределение точек (K, I_D) при комбинировании двух свидетельств в общем случае показано на Рис. 2. Нетрудно видеть, что с увеличением значения меры конфликта среднее значение индекса уменьшения незнания уменьшается, оставаясь положительным. При этом величина разброса слабо возрастает.

Из Рис. 3 видно, что в общем случае комбинирования свидетельств вероятность неотрицательности индекса уменьшения незнания остается большей 0,5 для всех значений меры конфликта.

2. Уточняющие свидетельства.

Совсем другой характер распределения меры конфликта и зависимости между мерой конфликта и индексом уменьшения незнания будет при комбинировании уточняющих свидетельств. На Рис. 4 приведена гистограмма распределения меры конфликта K в этом случае. Гистограмма показывает, что чаще всего встречаются уточняющие свидетельства со средней конфликтностью.

Из Рис. 5 видно, что в случае комбинирования уточняющих свидетельств с увеличением значения меры конфликта среднее значение индекса уменьшения незнания уменьшается и будет преимущественно отрицательным. При этом величина разброса резко возрастает.

Из Рис. 6 видно, что в случае комбинирования уточняющих свидетельств вероятность неотрицательности индекса уменьшения незнания будет меньше 0,5 для большинства значений меры конфликта.

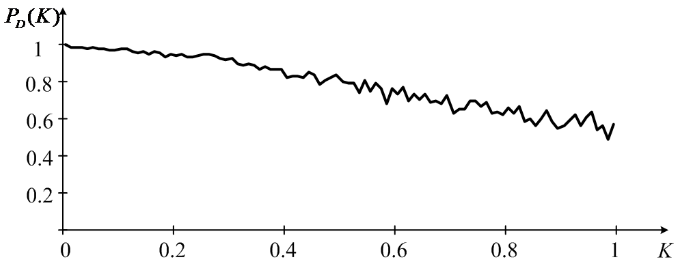

Рис. 3: Зависимость оценки вероятности $P_D(K) = P\{I_D > 0|K\}$ от конфликта в общем случае комбинирования

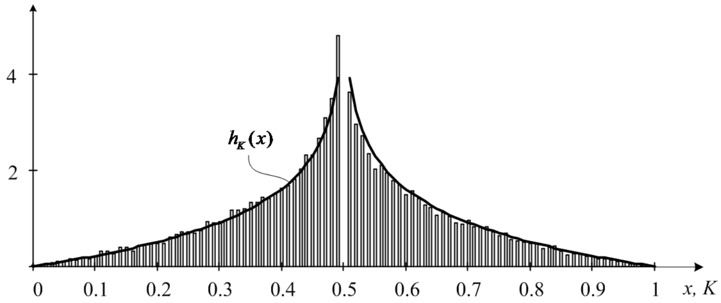

Рис. 4: Гистограмма и функция плотности конфликта в случае уточняющих свидетельств

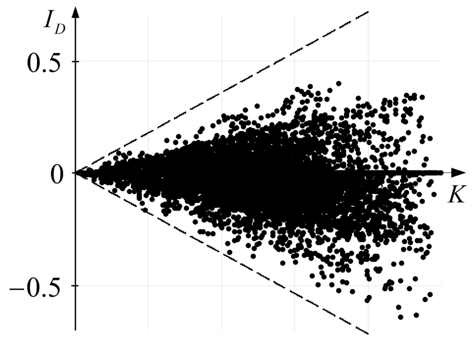

Рис. 5: Распределение точек (K, I_D) при комбинировании двух уточняющих свидетельств

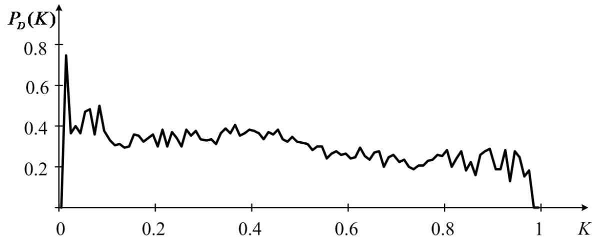

Рис. 6: Зависимость оценки вероятности $P_D(K)=P\{I_D>0|K\}$ от конфликта при комбинировании уточняющих свидетельств

4.2 Аналитическое исследование

Найдем сначала теоретические верхнюю и нижнюю оценки зависимости между мерой конфликтом и индексом уменьшения незнания.

Рассмотрим индекс уменьшения незнания $I_R(g_1, g_2)$ в случае комбинирования свидетельств по правилу R при условии, что мера конфликта $c(g_1, g_2) = K$, $g_1, g_2 \in Bel(X)$, $|X| = n$. Тогда величины

$$\bar{I}_R^{(n)}(K) = \sup\{I_R(g_1, g_2) : c(g_1, g_2) = K\}, \underline{I}_R^{(n)}(K) = \inf\{I_R(g_1, g_2) : c(g_1, g_2) = K\}$$

характеризуют верхнюю и нижнюю оценки уменьшения индекса незнания в зависимости от значения меры конфликта. Для величин $\bar{I}_D^{(n)}(K)$ и $\underline{I}_R^{(n)}(K)$ в случае комбинирования уточняющих свидетельств справедливо следующее утверждение.

Предложение 5. Пусть $F(g_1) = (\mathcal{A}(g_1), m_{g_1})$ и $F(g_2) = (\mathcal{A}(g_2), m_{g_2})$ два уточняющих тела свидетельств на X, связанные с функциями доверия $g_1, g_2 \in Bel(X)$, $|X| = n$, соответственно, причем $|\mathcal{A}(g_1)| = |\mathcal{A}(g_2)| = 2$. В случае комбинирования этих свидетельств по правилу Демпстера, вычислении индекса уменьшения незнания с помощью обобщенной меры Хартли, а меры конфликта по формуле (2), справедливы оценки $\bar{I}_D^{(n)}(K) \le K\frac{\ln(n-1)}{\ln n}$ и $\underline{I}_R^{(n)}(K) \ge -K\frac{\ln(n-1)}{\ln n}$ для всех $K \in (0, 1)$.

Замечание 6. Вычислительные эксперименты показывают, что неравенство для верхней оценки в Предложении 5 является завышенным. В частности, можно показать, что для верхней оценки изменения индекса незнания для «полуконфликтной» ($K = 0.5$) ситуации уточняющих свидетельств верно $\bar{I}_D^{(n)}(0.5) \le \frac{1}{\ln n}\ln\left(\frac{\sqrt{1+4n}-1}{2}\right)$. По всей видимости, и в общем случае (из которого получается последняя доказанная оценка при $K = 0.5$) $\bar{I}_D^{(n)}(K) \le Ky$, где y – положительный корень уравнения $n^{Ky} + n^y = n$.

На Рис. 5 графики верхней и нижней оценок функций $\bar{I}_D^{(6)}(K)$ и $\underline{I}_R^{(6)}(K)$, приведенные в Предложении 5, отмечены пунктиром.

Найдем теперь теоретическое распределение меры конфликта в случае уточняющих свидетельств при условии, что базовые вероятностные назначения этих свидетельств распределены по равномерному закону.

Предложение 6. Пусть два уточняющих тела свидетельств имеют мощность два, а их базовые вероятностные назначения распределены по равномерному закону. Тогда, случайная величина – каноническая мера конфликта K, вычисляемая по формуле (2), имеет плотность распределения, равную $h_K(x) = -\ln|2x-1|$, $x \in [0, 1/2) \cup (1/2, 1]$.

На Рис. 4 график функции $h_K(x)$ отмечен сплошной линией.

Заключение

В теории Демпстера-Шейфера существует проблема выбора правил комбинирования свидетельств. Решение этой проблемы связано с анализом количества незнания и конфликта, содержащейся в информации, представленной свидетельствами. В статье были введены индекс уменьшения незнания и мера конфликта для вычисления количества незнания и конфликта свидетельств. Эти индекс и мера исследуются в случае комбинирования двух тел свидетельств, удовлетворяющих определенным условиям и для различных правил комбинирования. Показано, что большая корреляция между телами свидетельств является достаточным условием уменьшения незнания после применения правила комбинирования. Если мы имеем несколько тел свидетельств (например, имеется несколько источников информации), мы можем использовать введенные индекс и меру для оптимального выбора тел свидетельств для комбинирования. Кроме того, эти инструменты можно использовать и для обоснованного выбора правил комбинирования.

Работа выполнена в рамках Программы фундаментальных исследований НИУ ВШЭ в 2013 году. Кроме того, исследование частично поддержано грантом РФФИ №11-07-00591.

Список литературы

[1] Dempster A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. The Annals of Statistics, 28, 1967, pp. 325 – 339.

[2] Shafer G. A Mathematical Theory of Evidence. Princeton, N.J.: Princeton University Press, 1976.

[3] Zadeh L.A. Review of Books: A Mathematical Theory of Evidence. The AI Magazine, 5(3), 1984, pp. 81 – 83.

[4] Yager R. On the Dempster-Shafer Framework and New Combination Rules. Information Sciences, 41, 1987, pp. 93 – 137.

[5] Dubois D., Prade H. Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence, 4, 1988, pp. 244 – 264.

[6] Sun Q., Ye X.Q., Gu W.K. A new combination rules of evidence theory. Acta Electronica Sinica, 28(8), 2000, pp. 117 – 119.

[7] Xin G., Xiao Y., You H. An Improved Dempster-Shafer Algorithm for Resolving the Conflicting Evidences. International Journal of Information Technology, 11(12), 2005, pp. 68 – 75.

[8] Deqiang H., Chongzhao H., Yi Y. A modified evidence combination approach based on ambiguity measure. Proc. of the 11th International Conference on Information fusion, 2008, pp. 1 – 6.

[9] Броневич А.Г., Лепский А.Е. Аксиоматический подход к определению индекса неточности нечетких мер. Труды 2-го международного семинара «Интегрированные модели и мягкие вычисления в искусственном интеллекте», М.: Физматлит, 2003, pp. 127 – 130.

[10] Bronevich A., Lepskiy A. Measuring uncertainty with imprecision indices. Proc. of the Fifth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'07), Prague, Czech Republic, 2007, pp. 47 – 56.

[11] Martin A., Jousselme A.-L., Osswald C. Conflict measure for the discounting operation on belief functions. Proc. of the 11th International Conference on Information fusion, 2008, pp. 1 – 8.

[12] Liu Z., Dezert J., Pan Q. A new measure of dissimilarity between two basic belief assignments. hal-00488045, version 1 - 1 Jun 2010.

[13] Lepskiy A. Estimation of Conflict and Decreasing of Ignorance in Dempster-Shafer Theory. Procedia Computer Science, 17, 2013, pp. 1113 – 1120.

[14] Smets P. Belief functions and transferable belief model. http://ippserv.rug.ac.be.

[15] Sentz K., Ferson S. Combination of evidence in Dempster-Shafer theory. Report SAND 2002-0835, Sandia National Laboratories, 2002.

[16] Inagaki T. Interdependence between Safety-Control Policy and Multiple-Sensor Schemes Via Dempster-Shafer Theory. IEEE Transactions on Reliability, 40(2), 1991, pp. 182 – 188.

[17] Zhang L. Representation, independence, and combination of evidence in the Dempster-Shafer theory. Advances in the Dempster-Shafer Theory of Evidence. R.R. Yager, J. Kacprzyk and M. Fedrizzi., New York, John Wiley and Sons, Inc., 1994, pp. 51 – 69.

[18] Деза Е.И., Деза М.-М. Энциклопедический словарь расстояний. М.: Наука, 2008.

[19] Dubois D., Prade H. On the combination of evidence in various mathematical frameworks. Reliability Data Collection and Analysis. J. Flamm and T. Luisi. Brussels, ECSC, EEC, EAFC, 1992, pp. 213 – 241.