УДК 504.052

М.З. ЗГУРОВСКИЙ, А.А. БОЛДАК

СИСТЕМНОЕ СОГЛАСОВАНИЕ ДАННЫХ РАЗНОЙ ПРИРОДЫ В МУЛЬТИДИСЦИПЛИНАРНЫХ ИССЛЕДОВАНИЯХ

Ключевые слова: *мультидисциплинарный подход, слабоструктурированная задача, согласованность данных, энтропия, анализ угроз.*

ВВЕДЕНИЕ

Количественные и качественные данные, используемые в мультидисциплинарных исследованиях, характеризуются, как правило, различной природой, изменяются в различных диапазонах, представляются в различных системах измерения [1, 2]. В этих случаях возникает потребность приведения таких данных к единой форме представления, «выравнивания» их масштабов, системного взаимного согласования, агрегирования их в удобные и более общие показатели [3].

Часто мультидисциплинарные задачи, совместно использующие результаты количественных измерений и экспертного оценивания, рассматриваются как слабоструктурированные [4]. В таком случае согласование данных следует рассматривать в более широком смысле — как решение задач двух типов.

Первый тип задач связан с методологией количественной оценки согласованности данных разной природы, второй — с разработкой методов такого согласования, когда оценки согласованности используются в качестве критериев оптимизации.

Несмотря на то что для решения задач обоих типов могут использоваться методы многомерного статистического анализа [3, 5], экспертно-статистического моделирования [6], экспертного оценивания [7], общий подход к системному согласованию данных разной природы отсутствует, что существенно сужает возможности мультидисциплинарных исследований.

Цель настоящей статьи состоит в разработке общего подхода и набора специализированных методов для решения задач системного согласования данных разной природы в рамках мультидисциплинарных исследований.

ОБЩИЙ ПОДХОД К СИСТЕМНОМУ СОГЛАСОВАНИЮ ДАННЫХ РАЗНОЙ ПРИРОДЫ

При оценке надежности, достоверности или обобщении различных данных следует оперировать разными взглядами на одно и то же явление или объект. Очевидно, что эти взгляды могут быть отражением различных свойств или различных объектов. На основании анализа согласованности таких взглядов делается вывод о надежности или достоверности одного из них.

Прежде всего необходимо выяснить, в чем заключается суть различий в природе данных, а затем формализовать понятие взглядов на данные различной природы. При этом будем исходить из того, что формальные взгляды представляют собой абстрактные объекты, на основе которых можно построить метрические пространства с мерами, определяющими их близость. Такие метрические пространства могут быть использованы для решения оптимизационных задач, связанных с согласованием различных взглядов.

Формальное определение взгляда. В общем случае исследование некоторого явления связано с анализом и обработкой сведений о нем. Эти сведения представляют собой количественные и (или) качественные оценки свойств некоторой совокупности объектов $O = \{o_i\}$, $i = \overline{1, n}$, где o_i — значения номинальной шкалы, идентифицирующие различные объекты из представленной совокупности. Считается, что такие данные являются результатом отображений вида

$$O \xrightarrow{I_j} X^j, \quad j = \overline{1, m}, \tag{1}$$

где I_j, $j = \overline{1, m}$, — отображения, определенные на множестве объектов ($D(I_j) = O$), с областями значений, равными областям определения показателей X^j, т.е. $E(I_j) = D(X^j)$.

Семантика отображений I_j, $j = \overline{1, m}$, и отождествляемых с ними показателей X^j в соотношении (1) формулируется, исходя из целей исследования, и определяет объективное со держание (свойство оценивания), целевое назначение (предназначения этих оценок) и способ получения (данные могут быть результатом измерений, моделирования, экспертного оценивания). Способ получения данных определяет также типы шкал, в которых измерены показатели [8].

Как видно из табл. 1, понятие согласованности данных играет ключевую роль при формулировке многих задач, разнообразие которых определяется целью исследования и способом получения данных.

Таблица 1

Интерпретация оценки согласованности	**Данные исследования (взгляд 1) различными способоми**			**Данные исследования (взгляд 2) различными способоми**		
	измерение	экспертные оценки	моделирование	измерение	экспертные оценки	моделирование
Надежность данных	Объект $O(t)$			Объект $O(t + \Delta t)$		
Достоверность данных	Объект $O_1(t)$			Объект $O_2(t)$		
Достоверность экспертных оценок	Объект $O(t)$				Объект $O_e(t)$	
Достоверность модели	Объект $O(t)$					Объект $O_M(t)$
Согласованность экспертных оценок		Объект $O_{e1}(t)$			Объект $O_{e2}(t)$	
Согласованность моделей			Объект $O_{M1}(t)$			Объект $O_{M2}(t)$

Данные вида (1) могут быть представлены таблицей «объект—свойство» [2]:

$$X = (x_{i,j})_{i=1, j=1}^{n,m}, \tag{2}$$

в которой строка X_i соответствует набору значений, характеризующему свойства объекта o_i, а столбец X^j задает значения j-го показателя для всей выборки объектов. Такая таблица является одной из форм описания отношения, заданного в пространстве показателей $X \subseteq O \times E(X^1) \times \ldots \times E(X^m)$.

Под взглядом на явление или объект будем понимать подмножество $V \subseteq X$. Это определение данных и взгляда является достаточно общим, поскольку позволяет представить такие понятия как «выборка объектов», «временной ряд», «проекция свойств» и т.д.

В общем случае если на взгляд V не накладывать ограничений, то, исходя из этого, можно определить все типы задач анализа согласованности данных — их непротиворечивости (целостности), оцениваемой как «близость» двух или нескольких взглядов на одно и то же явление или объект. Понятие «близости», а также конкретного содержания взглядов формулируется в рамках определенных задач исследования.

Построение метрических пространств для оценки согласованности данных. Наибольший интерес представляют взгляды двух типов: 1) выборка объектов, которая задает горизонтальные сечения таблицы (2), соответствующие множествам выбранных объектов; 2) проекция свойств (вертикальные сечения таблицы (2)), задающая значения используемых в ней показателей для всех объектов множества O.

В связи с этим элементарной выборкой будем называть выборку одного объекта $V_i \equiv X_i$, а элементарной проекцией — проекцию $V^j \equiv X^j$, которая задает значения одного показателя. Обозначим $\widetilde{V}$ выборку или проекцию, а $\widetilde{X}$ — элементарную выборку или проекцию.

Очевидно, что любая выборка или проекция представима в виде $\widetilde{V} \subseteq \bigcup_{i=1}^{|\widetilde{X}|} \widetilde{X}_i$.

Если на множестве $\widetilde{X} = \{\widetilde{X}_i\}$ задать σ-алгебру $\mathbb{S}$ [9] (тривиальная алгебра образуется как совокупность всех подмножеств исходных множеств), то можно определить метрическое пространство $(\widetilde{X}, \mathbb{S}, \mu)$, в котором конечно-аддитивная мера μ оценивает «близость» выборок или проекций.

Такой подход к определению мер μ не предполагает ограничений на природу данных, поэтому может использоваться в случае, когда данные измерены в номинальных шкалах. С другой стороны, возможности такого подхода ограничены лишь оценкой степени номинального совпадения различных взглядов.

Если мера близости взглядов выражает некоторые обобщенные свойства, необходимо рассматривать элементарные взгляды как элементы векторного пространства $E(X^1) \times \ldots \times E(X^m)$, на котором введена метрика.

В случае, когда удается определить пространство с нормой

$$||\widetilde{X}_i|| = \left(\sum_{j=1}^{|\widetilde{X}|} w_j (x_{i,j})^p \right)^{1/p}, \tag{3}$$

где w_j — весовые коэффициенты объектов или показателей, появляется возможность интегральной оценки взглядов с помощью нормы $||\widetilde{V}_i|| = \left(\sum_{\widetilde{X}_k \in \widetilde{V}_i} ||\widetilde{X}_k||^p \right)^{1/p}$, а также близости таких взглядов:

$$||\widetilde{V}_i|| - ||\widetilde{V}_j|| \le d(\widetilde{V}_i, \widetilde{V}_j) = ||\widetilde{V}_i - \widetilde{V}_j|| \le ||\widetilde{V}_i|| + ||\widetilde{V}_j||. \tag{4}$$

При $p = 2$ имеем евклидово пространство [9], в котором можно определить понятие ортогональности (независимости) взглядов.

Следует заметить, что определение расстояния в соотношении (4) может варьироваться в зависимости от целей исследования. Например, вместо $||\widetilde{V}_i - \widetilde{V}_j||$ часто используют норму $||\overline{V}_{i,j}||$, где $\overline{V}_{i,j}$ — некоторое усреднение (среднее или медиана) [10], найденное как решение задачи $\arg\min_{\overline{V} \in \mathbb{S}} (d(\overline{V}_{i,j}, \widetilde{V}_i) + d(\overline{V}_{i,j}, \widetilde{V}_j))$.

Могут быть использованы и другие нормы, наиболее точно отражающие взгляды, которые считаются близкими.

Индивидуальная нормировка данных как способ их системного согласования. Определение метрического пространства с нормой (3) предусматривает, что $X^j \in R,\ j=\overline{1,m}$, где R — поле действительных чисел [9]. Как отмечалось выше, показатели X^j могут отличаться по способу получения данных и, следовательно, типом применяемой шкалы, которая определяет свойства множества значений X^j. Если это условие не выполняется, то необходимо найти преобразования $C_j : X^j \xrightarrow{C_j} R$, называемые нормировками.

Для того чтобы сохранить отношения порядка, определенные на X^j, нормировки C_j должны быть монотонными, т.е. такими, для которых выполняется условие

$$x_{k,j} \prec x_{l,j} \Leftrightarrow C_j(x_{k,j}) \le C_j(x_{l,j});\quad k,l \in [1,n],\ \ k \ne l .$$

В случае, когда показатели X^j измерены в интервальных шкалах, можно использовать линейную нормировку

$$C_{\text{norm}}(x_{i,j}) = \frac{x_{i,j} - a}{b}, \tag{5}$$

где $x_{i,j}$ — значение из таблицы (2); a — параметр, задающий смещение; b — параметр, определяющий масштаб нормировки.

Очевидно, что нормированные значения являются безразмерными величинами, поскольку параметры a и b имеют ту же размерность, что и показатель X^j. Масштаб и смещение нормировки выбирают так, чтобы выполнялось условие $C_{\text{norm}}(x_{i,j}) \in [0,1]$ или $C_{\text{norm}}(x_{i,j}) \in [-1,1]$. Обычно для вычисления значений параметров a и b используют выражения

$$a = \min_{i=\overline{1,n}}(x_{i,j}),\ \ b = \max_{i=\overline{1,n}}(x_{i,j}) - \min_{i=\overline{1,n}}(x_{i,j}) \tag{6}$$

или

$$a = \overline{X}^j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j},\ \ b = \sigma(X^j) = \sqrt{\frac{\sum_{i=1}^{n}(x_{i,j} - \overline{X}^j)^2}{n}}, \tag{7}$$

где $\overline{X}^j$ — среднее значение, $\sigma(X^j)$ — стандартное отклонение показателя X^j, определяемое на выборке объектов O.

В случае использования соотношений (6) имеем $C_{\text{norm}}(X^j) \in [0,1]$, а с помощью соотношения (7) получим $\overline{C_{\text{norm}}(X^j)} = 0$ и $\sigma(C_{\text{norm}}(X^j)) = 1$. Соотношения (7) позволяют получить стандартизованные переменные [8], которые унифицированы по форме представления.

Линейные нормировки обладают свойством сохранять информацию, поэтому преобразованные показатели квалифицируются, как и исходные, статистическими характеристиками и приведены к единой форме представления данных.

Аналогично можно использовать нелинейные монотонные нормировки. Их применение целесообразно, когда с помощью показателей X^j необходимо оценить степень достижения некоторой интегральной цели, определяемой суммированием всех значений показателей. В таком случае может быть использовано соотношение

$$C_{\text{norm}}(x_{i,j}) = \left(1 - e^{\frac{a - x_{i,j}}{b}}\right)^{-1}, \tag{8}$$

задающее логистическую кривую [11], в которой параметр a определяет смещение, а b — масштаб кривой. Параметры a и b задаются так, чтобы соотношение (8) отображало степень достижения цели при значении показателя $x_{i,j}$. Очевидно, что получение такой оценки может быть слабо формализуемым. В этом случае параметры логистической кривой можно получить экспертным оцениванием, например методом непосредственного количественного оценивания [12]. Для этих целей используют лингвистическую переменную, которая определяется следующим образом.

Согласно [13] лингвистическая переменная формально определяется как пятерка $\{s, T(s), X, G, M\}$, где s — имя переменной; $T(s)$ — множество лингвистических значений переменной s, каждое из которых является нечетким множеством, определенным на множестве X; G — синтаксическое правило для образования значений переменной s; M — семантическое правило, связывающее каждое значение X с его семантикой $t(s) \in T(s)$.

В рассматриваемом нами случае $s =$ «Степень достижения цели», $T(s) =$ $=$ {«Низкая», «Высокая»}, X — область определения одного из показателей, G не определяется, а функции принадлежности M строятся на основе экспертных оценок.

Показатель X^j является областью определения функций принадлежности $\mu_s(x)$, для которых справедливо равенство $\mu_{\text{Низкая}}(x) = 1 - \mu_{\text{Высокая}}(x)$. Иными словами, для каждой пары «показатель–цель» достаточно определить только одну функцию принадлежности.

В литературе известны несколько способов моделирования функций принадлежности $\mu(x)$. Так, в работе [14] определяются π-, y-, t-, L-классы таких параметрических моделей, но, к сожалению, они не всегда удобны для практического экспертного оценивания параметров функции принадлежности. Функция $\mu(x)$ должна быть интуитивно понятной, иметь минимальное количество параметров, определяться на всем множестве X^j, принимать значения из интервала $[0, 1]$, быть монотонной и учитывать степень уверенности эксперта в точности его оценки.

Этими свойствами обладает нелинейная нормировка (8). При оценке ее параметров будем исходить из следующих допущений.

Во-первых, эксперт может задать значения показателя, которые являются необходимыми ($p_{\min}$) и достаточными ($p_{\max}$) для достижения цели.

Во-вторых, оценки $p_{\min}$ и $p_{\max}$ даются экспертом, исходя из его опыта, основанного на практической деятельности, когда он с одинаковой частотой сталкивается с фактами, соответствующими всему диапазону значений $[p_{\min}, p_{\max}]$. Это означает, что априорное распределение значений показателя можно считать равномерным в интервале $[p_{\min}, p_{\max}]$. Тогда параметры модели (8) могут быть определены следующим образом:

$$a = p_{\min} + \frac{p_{\max} - p_{\min}}{2}, \quad b = \frac{p_{\max} - p_{\min}}{\sqrt{12}}. \tag{9}$$

В соотношении (9) параметры a и b определяются соответственно как математическое ожидание и стандартное отклонение для случайной величины, равномерно распределенной в интервале $[p_{\min}, p_{\max}]$.

Применение оценок (8) с параметрами (9) позволяет перейти от описания объектов в виде

$$O = \{o_i \to < x_{i,1}, x_{i,2}, \ldots, x_{i,m} >;\ x_{i,j} \in X^j,\ \ j = \overline{1, m}\}, \tag{10}$$

где $X^1, X^2, \ldots, X^m$ — области значений показателей, к описанию вида

$$O = \{o_i \to < d_{i,1}, d_{i,2}, \ldots, d_{i,m} >;\ d_{i,1}, d_{i,2}, \ldots, d_{i,m} \in [0,1]\}. \tag{11}$$

В отличие от (11) в соотношении (10) области значений показателей X^j имеют разную природу, измеряются в разных шкалах и масштабах, а в (11) они отображены в безразмерные величины степени достижения цели из интервала $[0, 1]$. Это позволяет решить задачу согласования данных различной природы при комплексном оценивании степени достижения цели.

Частные оценки $\{d_{i,j}, j = \overline{1, m}\}$ удовлетворяют требованиям к полю, на котором определяется метрическое векторное пространство. Следовательно, они могут быть обобщены с помощью нормы вида (3).

Определение энтропийных мер отклонения для статистических распределений. Часто значения показателей X^j рассматриваются как статистические величины, закон распределения которых можно задать с помощью функции плотности распределения $f(x) = \dfrac{dF(x)}{dx}$, где $F(x) = P(X^j \le x)$ — интегральная функция распределения. Если показатель X^j является дискретной величиной, принимающей значения $\{x_l^j, \ l = \overline{1, k}\}$, функция $f(x)$ может быть задана вектором

$$P = \left\langle P(X^j = x_1^j), P(X^j = x_2^j), \ldots, P(X^j = x_k^j) \right\rangle = \langle p_1, p_2, \ldots, p_k \rangle \qquad (12)$$

при условии, что $p_l \ge 0, \ l = \overline{1, k}$, и $\sum_{l=1}^{k} p_l = 1.$

Закон распределения дискретной величины, заданный в виде (12), назовем кратко распределением.

В качестве оценки отклонения статистических распределений используют дивергенцию Кульбака–Лейблера [16]

$$D_{KL}(P \| Q) = \sum_{i=1}^{n} p_i(x) \log_2 \left(\frac{p_i(x)}{q_i(x)} \right), \qquad (13)$$

где $P = \langle p_1, p_2, \ldots, p_n \rangle$, $Q = \langle q_1, q_2, \ldots, q_n \rangle$ — распределения дискретной случайной переменной X, принимающей значения $\{x_1, x_2, \ldots, x_n\}$.

В работе [17] установлено, что $D_{KL}(P, Q) \ge 0$ (равенство достигается только в случае, когда $P \equiv Q$). Нетрудно показать, что

$$D_{KL}(P \| Q) = H(P, Q) - H(P), \qquad (14)$$

где $H(P, Q) = \sum_{i=1}^{n} p_i(x) \log_2 q_i(x)$ является перекрестной энтропией [18], а $H(P)$ — энтропией Шеннона [19], определяемой как

$$H(P) = \sum_{i=1}^{n} p_i(x) \log_2 p_i(x)\,.$$

Соотношение (14) устанавливает связь между информационными свойствами статистических переменных и отклонением их распределений от равномерного закона

$$H(X) = \log_2 n - D_{KL}(X \| X_U),$$

где X_U соответствует случайной величине, равномерно распределенной в диапазоне значений переменной X, принимающей значения $x_i, \ i = \overline{1, n}.$

Несмотря на то, что оценка (13) является неотрицательной и аддитивной [16, 17], она не может быть использована в качестве меры, так как $D_{KL}(P \| Q) \ne$ $\ne D_{KL}(Q \| P)$. Для получения симметричной оценки можно использовать соотношение

$$J(P, Q) = D_{KL}(P \| Q) + D_{KL}(P \| Q) = \sum_{i=1}^{n} (p_i - q_i) \log_2 \left(\frac{p_i}{q_i} \right). \qquad (15)$$

Оценки (13) и (15) существуют только тогда, когда $q_i > 0,\ i = \overline{1, n}$. Это справедливо для абсолютно непрерывных распределений Q относительно распределения P [16]. К сожалению, при решении практических задач часто возникает необходимость сравнения вырожденных распределений. В работе [20] предложены оценка вида

$$K(P,Q) = \sum_{i=1}^{n} p_i \log_2 \left(\frac{2p_i}{p_i + q_i} \right)$$

и соответствующая ей мера отклонения распределений

$$L(P,Q) = K(P,Q) + K(Q,P). \tag{16}$$

В работе [21] установлены верхняя граница меры $L(P,Q)$,

$$L(P,Q) \le \frac{1}{2} J(P,Q), \tag{17}$$

а также связь с энтропией Шеннона

$$L(P,Q) = 2H\left(\frac{P \oplus Q}{2} \right) - H(P) - H(Q), \tag{18}$$

где $\dfrac{P \oplus Q}{2} = \left\langle \dfrac{p_1 + q_1}{2}, \dfrac{p_2 + q_2}{2}, \ldots, \dfrac{p_n + q_n}{2} \right\rangle$ — усредненное распределение.

Меры (16) и (18) определены для любых распределений P и Q. На их основе можно разработать ряд оценок, имеющих практическую ценность при решении разнообразных задач.

Предлагаемый подход к построению таких критериев состоит в поиске граничных распределений: наилучшего $P_{\sup}$ и наихудшего $P_{\inf}$ с точки зрения решения конкретной прикладной задачи. Тогда критерий может быть определен следующим образом:

$$S(P) = \frac{L(P, P_{\inf}) - L(P_{\inf}, P_{\inf})}{L(P_{\sup}, P_{\inf}) - L(P_{\inf}, P_{\inf})}. \tag{19}$$

Если учесть, что $L(P_{\inf}, P_{\inf}) = 0$, то соотношение (19) принимает окончательный вид

$$S(P) = \frac{L(P, P_{\inf})}{L(P_{\sup}, P_{\inf})}. \tag{20}$$

Очевидно, что $S(P) \in \{0, 1\}$. Значение этого критерия может расцениваться как количественная оценка распределения P, выраженная в относительной шкале.

Пусть имеется множество распределений $\pi = \{P_j,\ j = \overline{1, m}\}$, а также определены граничные распределения $P_{\sup}$ и $P_{\inf}$. Тогда на множестве π может быть определен нестрогий порядок $\alpha \subseteq \pi \times \pi : \langle p_i, p_j \rangle \in \alpha \Leftrightarrow S(p_i) \ge S(p_j)$, т.е. решены оптимизационные задачи вида $\max\limits_{P \in \pi}(S(P))$, $\arg\max\limits_{P \in \pi}(S(P))$.

ПРАКТИЧЕСКИЕ ПРИЛОЖЕНИЯ

Рассмотрим особенности практического приложения представленного подхода на примере оценивания влияния совокупности угроз устойчивому развитию Украины [22]. Решение этой задачи связано с анализом двух категорий явлений.

С одной стороны, исследуемые процессы, охватывающие экономическую, социальную и экологическую сферы, имеют свойства, определяющие их желательное развитие (приоритеты безопасности). С другой стороны, в этих сферах можно выделить объекты или процессы, оказывающие негативное влияние на приоритеты безопасности (источники угроз), которые по своему объективному содержанию могут быть обобщены в рамках единого понятия угрозы.

Поэтому системное оценивание угроз предполагает получение интегрированного показателя уровня опасности (уровня уязвимости), количественно оценивающего негативное влияние совокупности угроз на выбранный приоритет безопасности. Этот интегрированный показатель должен быть построен на основе оценки совокупного негативного влияния угрозы, рассматриваемой как комплексное воздействие источников угроз, близких по своему объективному содержанию.

Следует заметить, что оценки уровня опасности имеют как объективную, так и субъективную составляющие. Одно и то же лицо в разных ситуациях может по-разному оценивать степень одного и того же негативного воздействия на один и тот же приоритет в зависимости от условно принятой этим лицом границы допустимого воздействия. Кроме того, могут существовать объективные пороги такого допустимого воздействия (например, влияние радиации на здоровье человека). Количественное оценивание уровня опасности должно учитывать обе эти составляющие.

Очевидно, что данные, количественно оценивающие проявление источников угроз, имеют разное объективное содержание, целевое назначение, а также способ получения. Это требует решения задач, согласования данных.

При системном оценивании угроз можно применить согласование целевого назначения показателей источников угроз с использованием индивидуальной нелинейной нормировки данных (8), параметры которой определяются согласно (9). С этой целью введем лингвистическую переменную s = «Уровень опасности» с областью значений $T(s)=$ {«Низкий», «Высокий»}.

В табл. 2 представлены перечень источников угроз приоритетам устойчивого развития Украины [22] и значения безопасного $p_{\min}$ (опасного $p_{\max}$) уровня для вычисления параметров функций принадлежности по формулам (9). Как видим, для одного и того же источника угроз и различных приоритетов устойчивого развития значения $p_{\min}$ и $p_{\max}$ могут не совпадать, что объясняется различной степенью влияния источников угроз на различные приоритеты. Более того, такое влияние может быть несущественным. В этом случае $p_{\min}$ и $p_{\max}$ не определяются.

Таким образом, уровень опасности j-го источника угрозы в отношении i-го приоритета безопасности

$$D_{i,j}=\{(x_j,\mu_{i,j}(x_j)|x_j\in X_j\}$$

является совокупностью значений количественной оценки источника угрозы x_j и значений уровня опасности $\mu_{i,j}(x_j)$, которые задаются с помощью функции $X_j \xrightarrow{\mu_{i,j}} [0,1]$.

Следует заметить, что для интегрированной оценки уровня опасности источников угроз использование операций над нечеткими множествами невозможно, так как в данном случае аргументами этих операций являются функции с различными областями определения [23].

В общей постановке задача интегрированной оценки заключается в формировании отображения $D_1\times D_2\times\ldots\times D_m\to I_{1,2,\ldots,m}\in R$, $D_1,D_2,\ldots,D_m=[0,1]$, множества векторов $< d_1,d_2,\ldots,d_m >$ в действительные числа, где $I_{1,2,\ldots,m}$ — значения нормы в пространстве $D_1\times D_2\times\ldots\times D_m$. Если учесть, что $D_1,D_2,\ldots,D_m=[0,1]$, то можно применить одну из Т-норм [24, 25], но в этом случае необходимо, чтобы $I_{1,2,\ldots,m}\in[0,1]$.

Предлагается интегрированной оценкой уровня опасности считать норму (3)

$$I_{1,2,\ldots,m}=||d||_p=\left(\sum_{i=1}^{m}w_i|d_i|^p\right)^{1/p}, \tag{21}$$

в которой $w_i\ge 0$, $\sum_{i=1}^{m}w_i=1$ — весовые коэффициенты, задающие значимость источников угроз. В качестве параметра p предлагается выбрать значение, равное количеству составляющих m.

Таблица 2

Код источника угрозы	Источник угрозы	Значения $p_{\min}$ ($p_{\max}$) уровня для приоритетов устойчивого развития			
		экономическое благосостояние (приоритет 1)	экологическая устойчивость (приоритет 2)	развитие гражданского общества (приоритет 3)	демографическое развитие (в том числе здоровье) (приоритет 4)
1	2	3	4	5	6
1	**Угроза «Демографический спад»**				
1.1	Общий коэффициент заключения браков	–	–	–	9.8(4.4)
1.2	Общий коэффициент разводов	–	–	–	0.15(5)
1.3	Демографическая нагрузка на население трудоспособного возраста	700(800)	–	–	700(800)
1.4	Общий коэффициент рождаемости	12(10)	–	–	12(8)
1.5	Общий коэффициент смертности	–	–	–	10(20)
2	**Угроза «Коррупция»**				
2.1	Вымогательство	5(30)	–	5(25)	–
2.2	Использование «связей»	5(30)	–	5(25)	–
2.3	Добровольные взятки	5(30)	–	5(25)	–
3	**Угроза «Социальная незрелость»**				
3.1	Уровень общего образования	0.4(0.2)	0.5(0.2)	0.3(0.1)	–
3.2	Уровень оплаты населением жилищно-коммунальных услуг	95(90)	–	95(90)	–
3.3	Уровень травматизма, связанного с производством	–	–	1(5)	–
3.4	Коэффициент смертности населения от преднамеренных увечий	–	–	2(20)	20(45)
3.5	Коэффициент преступности	–	–	250(1000)	–
3.6	Активность на выборах	–	–	70(50)	–
3.7	Уровень патриотизма	–	–	77(9)	–
4	**Угроза «Политическая нестабильность»**				
4.1	Уровень доверия к государственным институтам: местные органы власти	–	–	50(15)	–
4.2	Уровень пассивности населения относительно потенциального участия в общественных и политических организациях	–	–	60(90)	–
4.3	Оценка критичности политической ситуации в стране	–	–	15(25)	–
4.4	Готовность к участию в акциях протеста	–	–	20(75)	–
5	**Угроза «Снижение здоровья людей»**				
5.1	Численность ВИЧ-инфицированных по регионам	–	–	–	0.01(0.1)
5.2	Заболеваемость населения активным туберкулезом по регионам	–	–	–	10(100)
5.3	Коэффициент смертности младенцев	–	–	–	5(15)
6	**Угроза «Загрязненность окружающей среды»**				
6.1	Качество водопроводной воды по химическим показателям	–	0(20)	–	0(20)
6.2	Качество водопроводной воды по бактериологическим показателям	–	0(10)	–	0(10)
6.3	Суммарное загрязнение атмосферного воздуха	–	5(7)	–	5(7)
7	**Угроза «Природные стихийные бедствия»**				
7.1	Подтопление	10(50)	10(50)	–	10(50)
7.2	Активные оползни	1(5)	1(5)	–	–
7.3	Сейсмическая активность	4(8)	4(9)	–	4(10)

1	2	3	4	5	6
8	**Угроза «Антропогенная перегруженность»**				
8.1	Застроенные земли	–	2(12)	–	–
8.2	Сельскохозяйственные угодья	–	30(85)	–	–
8.3	Природные источники воды	–	100(300)	–	–
8.4	Выбросы загрязняющих веществ	–	0(10)	–	–
8.5	Плотность выбросов загрязняющих веществ	–	10(50)	–	5(25)
8.6	Плотность выбросов парниковых газов	–	200(1000)	–	–
9	**Угроза «Техногенная опасность»**				
9.1	Степень износа основных фондов	40(60)	30(70)	–	40(60)
9.2	Население в зонах возможного химического заражения	–	20(50)	–	10(60)
9.3	Населенные пункты в зонах радиоактивного загрязнения	–	0(10)	–	0(10)
9.4	Населенные в зонах возможного радиоактивного заражения	–	20(50)	–	10(40)
10	**Угроза «Ресурсозависимость»**				
10.1	Доля земельных угодий, используемых сельскохозяйственными предприятиями и хозяйствами	0.3(0.7)	0.3(0.5)	–	–
10.2	Среднегодовая занятость населения	0.2(0.53)	–	0.4(0.7)	0.6(0.2)
10.3	Уровень энергетической обеспеченности промышленного производства	0.1(0.5)	0.5(0.1)	–	–
10.4	Степень активности внешнеэкономической деятельности	0.1(0.5)	–	–	–
11	**Угроза «Дисбаланс производственной инфраструктуры»**				
11.1	Инвестиционная привлекательность и инвестиционные возможности	0.1(0.5)	–	–	–
11.2	Обеспеченность основными фондами	0.1(0.6)	0.4(0.1)	–	–
11.3	Промышленно-сельскохозяйственный баланс	0.3(0.64)	0.3(0.64)	–	–
12	**Угроза «Производственно-потребительский дисбаланс»**				
12.1	Доходно-расходный баланс населения	0.25(0.4)	–	–	0.25(0.4)
12.2	Валовой региональный продукт на душу населения	0.4(0.7)	–	–	–
12.3	Спрос на рабочую силу	0.1(0.9)	–	–	0.8(0.69)

Следует заметить, что верхняя граница области значений полученной интегральной оценки степени опасности лишь асимптотически приближается к единице при $p \to \infty$, т.е. такая оценка не является Т-нормой в классическом понимании. С другой стороны, это позволяет выполнять сравнение значений уровня опасности угроз, близких к единице.

Интеграция уровня опасности совокупности угроз по отношению к приоритетам устойчивого развития может осуществляться также с использованием нормы (21). Тогда весовые коэффициенты w_i должны отображать степень важности этих приоритетов.

Анализ определенности экспертных оценок для уровня опасности угроз. Для определения весовых коэффициентов в формуле (21) используются различные методы. Когда эти коэффициенты отображают степень важности приоритетов устойчивого развития, которая в большей степени носит субъективный характер, предпочтительным является применение методов экспертного оценивания. Если такие качества экспертов, как их уверенность в правильности, однозначности (определенности) своих выводов влияют на результаты экспертного оценивания, следует применять опосредованные оценки самих экспертов. Это основано на анализе энтропии, привносимой экспертами в результаты оценивания.

После нахождения экспертным путем параметров (9) функций принадлежности (8) и получения оценок уровня опасности угроз можно вычислить степень определенности такой экспертизы. При этом критерий определенности экспертной оценки S имеет вид (20), а его значения используются в интегрированной оценке (21). Для этого необходимо определить наилучшее $P_{\sup}(X_i)$ и наихудшее $P_{\inf}(X_i)$ распределение значений уровня опасности для переменной X_i в смысле определенности экспертной оценки.

Наибольшей неопределенности экспертной оценки соответствует случай, когда эксперт независимо от значения x_i характеризует его как средний уровень опасности, т.е. $\forall x_i \in X_i : \mu(x_i) = 0{,}5$. Тогда для m интервалов группирования данных имеем вырожденное распределение

$$P_{\inf}(X_i, m) = \begin{cases} \langle 0,0, \ldots, p_{k+1}, \ldots, 0,0 \rangle, & p_k = 1, \ m = 2k+1, \ k \in N, \\ \langle 0,0, \ldots, p_k, p_{k+1}, \ldots, 0,0 \rangle, & p_k = p_{k+1} = 0{,}5; \ m = 2k, \ k \in N. \end{cases}$$

Наименьшая неопределенность экспертной оценки соответствует случаю, когда эксперт относит значения x_i к одному из двух классов: «Низкий уровень опасности» и «Высокий уровень опасности», используя значение параметра смещения a в выражении (9) для функции принадлежности (8) в качестве разделяющего значения. В этом случае наилучшее распределение примет вид

$$P_{\sup}(X_i) = \left\langle \frac{m_0}{m}, 0, \ldots, 0, \frac{m_1}{m} \right\rangle,$$

где m_0 и m_1 — число значений x_i, отнесенных к классу «Низкий уровень опасности» и «Высокий уровень опасности» соответственно.

На рис. 1 и рис. 2 даны графики поведения критерия S при варьировании параметров a и b.

На рис. 1 наблюдается минимальное значение S вблизи значения 0. Это объясняется тем, что функция $\mu(x_i)$ близка к линейной зависимости (средний сегмент логистической кривой), что соответствует наибольшей степени неопределенности.

Как видно из рис. 2, степень определенности выше среднего уровня наблюдается в том случае, когда значение параметра b составляет величину, меньшую, чем 20–30% стандартного отклонения σ.

Рассчитанные значения степени определенности экспертных оценок для параметров функций принадлежности представлены диаграммой на рис. 3. Эти оценки были использованы для нахождения весовых коэффициентов при вычислении уровня опасности угроз приоритетам устойчивого развития для Украины.

В табл. 3 сведены усредненные по регионам Украины оценки уровня опасности угроз приоритетам устойчивого развития,

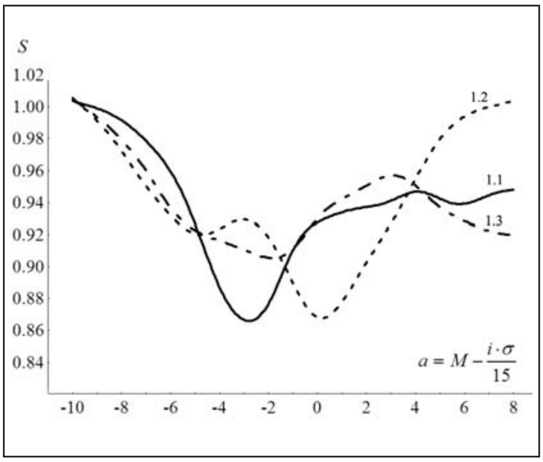

Рис. 1

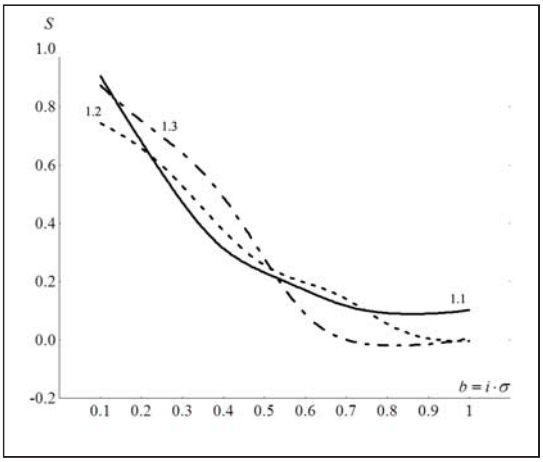

Рис. 2

а также определен масштаб проявления влияния этих угроз. Как видно из таблицы, в среднем для Украины можно определить наиболее значимые по уровню опасности угрозы.

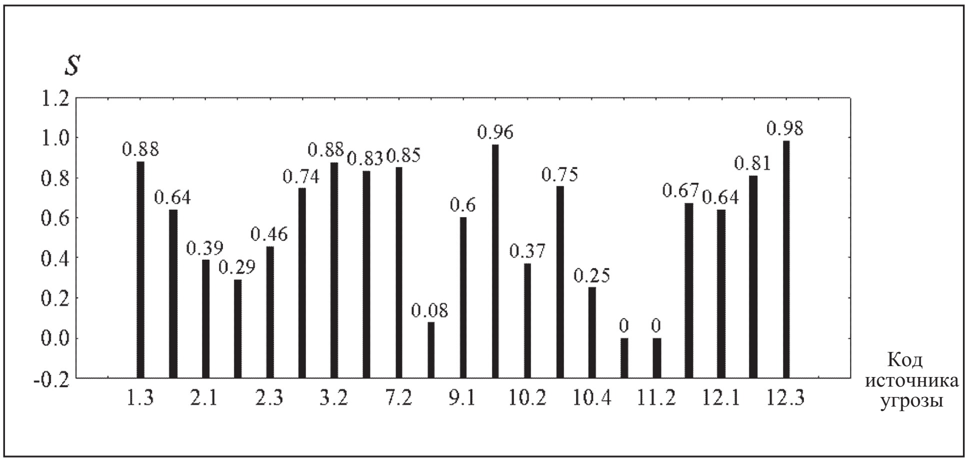

Рис. 3

Таблица 3

Код угрозы	Среднее значение уровня опасности угроз по приоритетам				Масштаб проявления влияния угроз по приоритетам			
	1	2	3	4	1	2	3	4
1	0,51	0,00	0,00	**0,78**	0,33	0,00	0,00	**0,67**
2	**0,67**	0,00	**0,70**	0,00	0,37	0,00	0,37	0,00
3	0,31	0,00	**0,85**	0,21	0,07	0,00	**0,78**	0,00
4	0,00	0,00	**0,91**	0,00	0,00	0,00	**0,85**	0,00
5	0,00	0,00	0,00	**0,82**	0,00	0,00	0,00	**0,67**
6	0,00	**0,83**	0,00	**0,83**	0,00	**0,67**	0,00	**0,67**
7	**0,71**	**0,68**	0,00	0,50	0,44	0,44	0,00	0,19
8	0,00	**0,76**	0,00	0,34	0,00	0,63	0,00	0,19
9	**0,64**	**0,88**	0,00	**0,91**	0,44	**0,81**	0,00	**0,85**
10	**0,86**	**0,55**	0,31	0,38	**0,85**	0,37	0,00	0,04
11	**0,87**	**0,59**	0,00	0,00	**0,96**	0,30	0,00	0,00
12	**0,99**	0,00	0,35	0,35	**1,00**	0,00	0,00	0,15

Так, для экономической сферы (приоритет 1) наиболее опасными являются следующие угрозы: «Производственно-потребительский дисбаланс», «Дисбаланс производственной инфраструктуры», «Ресурсозависимость»; для экологического состояния страны — «Природные стихийные бедствия» и «Техногенная опасность»; для социальной сферы — «Коррупция».

С точки зрения экологической устойчивости (приоритет 2) наиболее опасными являются угрозы «Загрязненность окружающей среды», «Природные стихийные бедствия», «Антропогенная перегруженность», «Техногенная опасность», «Ресурсозависимость», «Дисбаланс производственной инфраструктуры».

На развитие гражданского общества (приоритет 3) большое негативное влияние оказывают угрозы социального характера: «Коррупция», «Социальная незрелость» и «Политическая нестабильность».

Демографическое развитие и здоровье людей (приоритет 4) подвергается ощутимому влиянию социальных («Демографический спад», «Снижение здоровья людей») и экологических («Загрязненность окружающей среды», «Техногенная опасность») угроз.

Масштаб проявления влияния угроз определяется как доля от общего числа тех регионов Украины, для которых уровень опасности угроз выше среднего. Исходя из этого можно выделить общегосударственный и региональный масштабы проявления угроз. Так, например, существенное влияние «Дисбаланса производственной инфраструктуры» на экономическое благосостояние имеет общегосударственный масштаб, а влияние этой же угрозы на экологическую устойчивость является существенным, но проявляется в региональном масштабе.

Информация, предоставленная в табл. 2 и табл. 3, может использоваться при принятии решений на государственном и региональном уровне, когда необходимо идентифицировать негативные факторы, оказывающие существенное влияние на процессы устойчивого развития.

ЗАКЛЮЧЕНИЕ

В настоящей статье рассмотрен подход к системному согласованию данных разной природы, используемых в мультидисциплинарных исследованиях, которые отличаются объективным содержанием, целевым назначением и способом получения.

На основе модели данных «объект–свойство» предложен подход, состоящий в определении согласованности данных как количественной оценки близости двух или более взглядов на одно и то же явление или объект. Рассмотрены особенности построения метрических пространств, используемых для оценки согласованности данных. Построены энтропийные меры отклонения статистических распределений для количественной оценки степени неопределенности экспертных оценок.

Систематизированы способы индивидуальной нормировки данных, которая рассматривается как способ их согласования. Введены лингвистические переменные и предложен способ задания функций принадлежности с помощью логистической кривой для решения слабоструктурированных задач согласования показателей в задачах комплексного оценивания степени достижения цели.

Теоретические результаты были использованы для решения задачи системного согласования данных при комплексном оценивании влияния совокупности угроз на устойчивое развитие регионов Украины. Практические результаты будут полезны для лиц, принимающих решения на региональном и государственном уровне.

СПИСОК ЛИТЕРАТУРЫ

1. Somervill M., Rapport D. Transdisciplinarity: recreating integrated knowledge. — Oxford (UK): EOLSS Publ. Co. Ltd., 2000 — 271 p.
2. Айвазян С.А. Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справ. изд. — М.: Финансы и статистика, 1985. — 487 с.
3. Аналіз сталого розвитку — глобальний і регіональний контексти / М.З.Згуровський, А.О. Болдак, С.В. Войтко та інш. — 2010. — Ч. 1. — (www.wdc.org.ua)
4. Newell A., Simon H. Human problem solving. Englewood Cliffs: (N J): — Prentice-Hall, 1972. — 920 p.
5. Померанцева Т.Н., Болдак А.А. Многомерный статистический анализ влияния глобальных угроз на безопасность стран мира // Кибернетика и системный анализ. — 2010. — № 2. — С. 37–48.
6. Згуровский М.З., Болдак А.А., Померанцева Т.Н. Анализ влияния глобальных угроз на устойчивое развитие стран и регионов мира с помощью байесовских сетей доверия // Там же. — 2010. — № 5. — С. 152–163.
7. Zgurovsky M.Z. System adjustment of various nature DATA for global modelling of sustainable development // Proc. of 22-nd Intern. CODATA Conf. 24–27 Oct., 2010, Cape Town, South Africa. — 27 p.
8. Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 607 с.

9. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. — 7-е изд. — М.: Физматлит, 2004. — 572 с.
10. Орлов А.И. Эконометрика. — М.: Экзамен, 2004. — 576 с.
11. Balakrishnan N. Handbook of the Logistic Distribution. — New York: Marcel Dekker, 1992. — 601 p.
12. Лисецкий Ю.М., Каревина Н.П. Об автоматизации экспертных оценок // Мат. машини і системи. — 2008. — № 1. — С. 151–162.
13. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. — М.: Мир, 1976. — 166 с.
14. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы / Пер. с пол. И.Д. Рудинского. — М.: Горячая линия — Телеком, 2004. — 452 с.
15. Trench W. Introduction to real analysis. Upper Saddle River (N.J.): Prentice Hall / Pearson Educ., 2002. — 504 p.
16. Kullback S., Leibler R. On information and sufficiency // Ann. Math. Statist. — 1951. — **22**, N 1. — P. 79–86.
17. Johnson R. Axiomatic characterization of the directed divergences and their linear combinations // IEEE Trans. Inform. Theory. — 1979. — **IT-25**, N 6. — P. 709–716.
18. Габидулин Э.М., Пилипчук Н.И. Лекции по теории информации. — М.: МФТИ, 2007. — 214 с.
19. Шеннон К. Работы по теории информации и кибернетике. — М.: Изд-во иностр. лит., 2002. — 836 с.
20. Lin J., Wong S. Approximation of discrete probability distributions based on a new divergence measure // Congressus Numerantiitm. — 1988. — **61**. — P. 75–80.
21. Lin J. Divergence measure based on the Shannon entropy // IEEE Trans. Inform. Theory. — 1991. — **37**, N 1. — P. 145–191.
22. Аналіз сталого розвитку — глобальний і регіональний контексти / М.З. Згуровський, А.О. Болдак, С.В. Войтко. — 2010. — Ч. 2. — (www.wdc.org.ua).
23. Батыршин И.З. Основные операции нечеткой логики и их обобщения. — Казань: Отечество, 2001. — 100 с.
24. Cignoli R., D'Ottaviano M., Mundici D. Algebraic foundations of many-valued Reasoning. — Dordrecht: Kluwer Acad. Publ., 2000. — 244 p.
25. Klement E., Mesiar R., Pap E. Triangular norms. — New York: Springer, 2010. — 406 p.
26. Шторм Р. Теория вероятностей. Математическая статистика. Статистический контроль качества. — М.: Мир, 1970. — 368 с.
27. Heinhold I., Gaede K.W. Ingeniur statistic. — München; Wien, Springler-Verlag, 1964. — 352 p.
28. Chou Y. Statistical analysis with business and economic applications. — Holt: Rinehart and Winston, 1975. — 794 p.

Поступила 25.04.2011