

В.В. Осипенко

Оценивание релевантности результатов в индуктивных процедурах системно-аналитических исследований

Рассмотрен один из ключевых вопросов системно-аналитических исследований – критерии выбора оптимальных решений. Даны определения критериев релевантности и корелевантности, сконструированных на базе матриц экспертных оценок. Приведена структура индуктивной процедуры системно-аналитических исследований, использующая принципы метода группового учета аргументов.

One of the key questions of the system-analytical investigations – the criteria of the choice of optimal solutions is considered. The relevance and corelevance criteria based on the expert values matrix are defined. A structure of the inductive procedure using the principles of the method of group handing is presented.

Розглянуто одне з ключових питань системно-аналітичних досліджень – критерії вибору оптимальних рішень. Дано визначення критеріїв релевантності та корелевантності, побудованих на базі матриць експертних оцінок. Наведено структуру індуктивної процедури системно-аналітичних досліджень із застосуванням основних принципів методу групового урахування аргументів.

Введение. Современная информационно-аналитическая деятельность и системно-аналитические исследования (САИ) – более широкие понятия, чем сфера информационного обслуживания, такого, например, как реферирование или написание аналитического обзора, библиографической справки и множества других типов документов информационно-аналитического сервиса (ИС) или (ИАС) компилятивного характера. Информационно-аналитический сервис получил развитие благодаря быстрорастущим требованиям практически во всех сферах занятости современного человека. Тем не менее, несмотря на очевидные принципиальные различия, САИ и ИС объединяют, по крайней мере, два обязательных главных элемента: вид результата – это некоторый «документ» и критерий соответствия: в ИАС результата запроса, в САИ – реальности в отношении исследуемого сложного объекта (процесса, явления, проблемы). В информационной деятельности такой документ даже получил специальное узаконенное название – «вторичный документ» [1].

Определение результата в САИ

Хотя понятие *результата* («документа») в САИ более размыто и гораздо сложнее, чем строгая дефиниция закона для «вторичного документа» в ИС, в дальнейшем исследовании необходимо использовать некоторое формальное его описание. При этом за формальной оболочкой понятия результата САИ желательно не потерять естественную смысловую нагрузку последнего. Исходя из этого, попыта-

емся дать следующее определение для результата САИ в устоявшемся уже современном его понимании.

Определение 1. *Результатом* $R^*(I_b^*)$ комплексного системно-аналитического исследования есть специфический документ $D\{R^*(I_b^*)\}$, в котором отображены результаты системного анализа сложного объекта (процесса, явления, проблемы), который базируется на сконструированном в процессе исследования оптимальном информационном базисе I_b^* , отвечает установленным требованиям, имеет информационно-рекомендательный характер, наделен определенным официальным статусом и уровнем доступа.

В длинный перечень требований заказчика чаще всего входит и необходимость разработки комплексных количественных прогнозов развития объекта исследования для принятия статусных решений с предвидением последствий от их реализации и т.п. Таким образом, САИ есть самодостаточный системно-аналитический *проект*, но, как частный случай, может быть также существенным дополнительным элементом в интеллектуальных системах поддержки принятия решений (ИСППР), а его результат – «документ» – может иметь вид, например, некоторого специальным образом организованного файла базы знаний такой системы. Кроме того, важно подчеркнуть, что выходной документ не обязательно может относиться к одной определенной предметной области (в отличие

от такого требования в экспертных системах, например), а охватывать несколько предметных областей, в том числе и являющихся смежными.

Следовательно, современное понятие результата САИ в широком смысле не принадлежит к традиционным информационным документам типа реферат, аналитический или библиографический обзор и другим, а намного шире и включает, в том числе и их как составные части.

Примерами результатов САИ могут быть:

- инновационно-инвестиционный проект развития предприятия, отрасли, страны;
- инвестиционный меморандум;
- комплексный стратегический план развития банковской, страховой или торговой сети;
- прогноз развития геополитической ситуации в Европе в среднесрочной (5–10 лет) или долгосрочной, т.е. на 10 и более лет, перспективе;
- прогноз влияния климатических изменений в среднесрочной (5–10 лет) или долгосрочной (10–20 и более лет) перспективе на развитие аграрного сектора европейских стран и т.п.

Очевидно, что определяющие ресурсы, необходимые для таких исследований (время, средства, информация, профессиональные высокооплачиваемые кадры и т.п.) весьма значительны.

Вторым из двух главных упомянутых элементов есть критерий (или система критериев) соответствия такого результата: вторичного документа – запросу (в ИС) или «реальности» в отношении исследуемого объекта (в САИ проблемного характера). Здесь центральное место занимает понятие *релевантности* (*relevance* – англ.: соответствие), одно из базовых понятий в системах поиска и анализа информации в современных информационно-аналитических, информационно-поисковых системах, в том числе *WEB*-ориентированных. В информационной деятельности сегодня в основном господствует определение релевантности как «соответствие результирующего документа информационного поиска запросу» [например, 2 и др.]. Исходя из этого, исследования в этом направлении в ос-

новном сводятся к разработке способов формализации мер подобия документов и соответствия выходного документа именно запросу, и в этом направлении имеется огромное количество публикаций.

Определение критерия релевантности в системно-аналитических исследованиях

В нашем случае понятие *релевантности* имеет принципиальное отличие от общепринятого понятия, используемого в информационно-поисковых системах определенной предметной направленности. Это отличие состоит в том, что в САИ необходимо оценивать не сходство результирующего документа с запросом, а соответствие результата информационно-аналитического исследования поставленным требованиям (заказчика, например) и, что более важно, *адекватности* результата системного анализа решаемой проблеме, или степень соответствия *реальности* анализа явлению, процессу и т.п. Исходя из уникальности и сложности объектов исследования, в работе [3] высказана идея рассматривать проект САИ как некоторый аналог многоэтапных алгоритмов метода группового учета аргументов (МГУА). Кроме того, там указаны аналогии и принципиальные отличия между проектом САИ и многоэтапным алгоритмом МГУА, а также высказаны возможные пути их преодоления. Ключевыми направлениями являются: определение и формализация понятия релевантности в САИ и определение принципиальной возможности применения формальных (количественных) аналогов внешних критериев оценивания качества моделей МГУА в проектах синтеза результатов сложных информационно-аналитических исследований, а отсюда и возможности применения методологии МГУА в САИ.

Для определения понятия релевантности с целью конструирования специальных внешних критериев оценивания результатов в комплексных системно-аналитических исследованиях, по мнению автора, достаточно адекватно и эффективно может быть использован *аппарат нечетких множеств* (фаззилогики). За 45 лет с момента создания теоретические и убедительные прикладные аспекты его описаны в мно-

гочисленной литературе, в том числе в [4]. Применение нечетких моделей и методов достаточно широко представлено и в современной отечественной литературе по интеллектуальным системам принятия решений [5]. Действительно, этот аппарат оказывается особенно полезным в случаях, когда процессы и объекты системных исследований слишком сложны для анализа с помощью общепринятых количественных методов или когда доступные источники информации интерпретируются на некачественном уровне, неточно или неопределенно, а процедура анализа не поддается строгому математическому описанию.

Введем понятие образа «целевого» (эталона) результата САИ – $R^0(I_b^0)$. Это – прямоугольная матрица $E = (e_{ij})$ размерности $n \times m$

$$E = (e_{ij}) = \begin{pmatrix} e_{11} & \dots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \dots & e_{mn} \end{pmatrix}, \quad (1)$$

i -я строка которой ($i = 1, 2, \dots, m$), отображает один из беспрекословных видов требований к «целевому» результату исследования с позиций экспертов верхнего уровня, а j -й столбец ($j = 1, 2, \dots, n$) – возможные градации оценок i -го элемента. Строкой может быть, например, один из необходимых разделов будущего документа, а элементами строки (ячейки) – экспертные оценки требований к его сути. Кроме того, эксперты все-таки допускают определенные отклонения будущего результата от «целевого идеала», но при этом оценивают подобные «несовершенства». Такая матрица формируется экспертным способом на стадии постановки проекта с использованием аппарата, например, фаззилогики, но это не предмет обсуждения в этой статье.

Введем следующие обозначения. Пусть $E(R^0(I_b^0))$ – матрица экспертных значений «целевого» результата $R^0(I_b^0)$ вида (1), т.е. все значения ее элементов полностью удовлетворяют заказчика исследования или результат $R^0(I_b^0)$ решает возникшую проблему «наилучшим» из возможных образом. Далее, пусть $W(R_k(I_b^j))$ –

матрица некоторого частного результата $R_k(I_b^j)$, отвечающего текущему информационному базису I_b^j на j -м этапе (шаге) системно-аналитического проекта; I_b^1 – заданный первичный информационный базис, а I_b^+ – некоторая целевая порция мониторинговой информации, которая должна дополнять уже имеющийся информационный ансамбль I_b^j с целью улучшения результата $R_k(I_b^j)$, приближая его к «целевому» $R^0(I_b^0)$; $R^*(I_b^*)$ – оптимальный результат, который может быть достигнут в рассматриваемом системном информационно-аналитическом проекте.

Обозначим через Δ_{rel}^2 матрицу

$$\Delta_{rel}^2 = \begin{pmatrix} \delta_{11}^2 & \dots & \delta_{1n}^2 \\ \vdots & \ddots & \vdots \\ \delta_{m1}^2 & \dots & \delta_{mn}^2 \end{pmatrix} = (\delta_{ij}^2), \quad (2)$$

$$i = 1, \dots, m, \quad j = 1, \dots, n,$$

элементы δ_{ij}^2 которой равны квадратам разностей элементов матриц $W(R_k(I_b^j))$ и $E(R^0(I_b^0))$ (обе матрицы имеют одинаковые размерности $n \times m$).

Определение 2. Под *системной релевантностью* (*Rel*) в системно-аналитическом исследовании будем понимать меру вида:

$$CR_{rel} = \|W(R_k(I_b^j)) - E(R^0(I_b^0))\|, \quad (3)$$

а критерием системной релевантности назовем выражение:

$$CR_{rel} = \left(\sum_{i=1}^m \sum_{j=1}^n (\delta_{ij}^2)_{WE} \right)^{\frac{1}{2}}. \quad (4)$$

Таким образом, выражение критерия системной релевантности в комплексных системно-аналитических исследованиях принял простой и понятный смысл.

Постановку задачи индуктивной процедуры САИ теперь можно сформулировать следующим образом: на основе заданного первичного информационного базиса I_b^1 и полученной в процессе исследования дополнительной целе-

вой информации $\{I_b^+\}$ необходимо синтезировать некоторое множество \mathfrak{R} результатов вида $R(I_b) \in \mathfrak{R}$ с целью нахождения оптимального результата по минимуму заданного критерия $CR_{rel}(\cdot)$.

Определение 3. Результат $R^*(I_b^*)$, для которого значение критерия (4) минимально, т.е.

$$R^*(I_b^*) = \arg \min_{R(I_b) \in \mathfrak{R}} CR_{rel} \{W(R_k(I_b^j)), E(R^0(I_b^0))\}, \quad (5)$$

называется *оптимальным результатом комплексного системно-аналитического исследования*.

Оптимальный результат $R^*(I_b^*)$ в дальнейшем – основа для создания окончательного структурированного согласно требованиям документа $D\{R^*(I_b^*)\}$.

Критерий CR_{rel} – интегральный (системный) критерий релевантности верхнего уровня и, соответственно, системная оценка качества результата САИ.

Критерий корелевантности в индуктивных процедурах САИ как аналог критерия минимума смещения МГУА

Рассмотрим индуктивную процедуру системно-аналитического исследования, базирующуюся на традиционной технологии анализа, несколько видоизменив последнюю. Поскольку в дальнейшем будет рассмотрена общая схема индуктивной технологии САИ, основанная на методологии метода группового учета аргументов, назовем такую схему «однотуповой». Схема этой технологии хорошо согласуется с традиционными процедурами информационно-аналитических исследований (за исключением критерия $\delta = \|W, E\|$ и, возможно, блока S4) и показана на рис. 1.

Обозначения на рис. 1 имеют следующий смысл: S1 – блок управления целенаправленной процедурой исследования (проектный блок); S2 – блок системного анализа, превращения информации и моделирования; S3 – блок синтеза результата $R_k(I_b^j)$ по базису I_b^j и формирова-

ния матрицы экспертных оценок его качества $W(R_k(I_b^j))$; S4 – блок информационного мониторинга внешнего информационного пространства и фильтрации информации; $U(I_b^+)$ – управляющее воздействие на блок S4 для получения I_b^+ (это может быть, например, некоторый листинг наиболее важной на данном этапе информации).

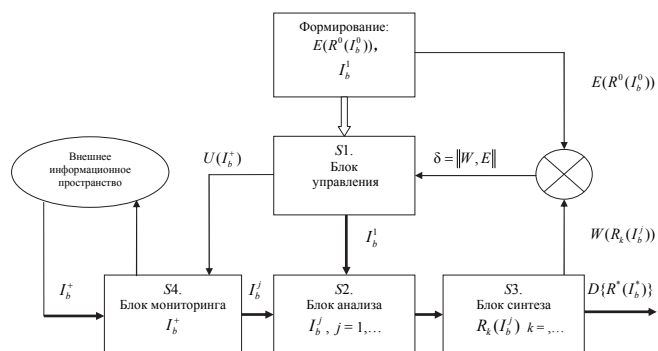


Рис. 1. Информационно-технологическая схема работы одной аналитической группы в САИ

Как видно из рис. 1, эта технология внешним дополнением в классическом понимании его в теории МГУА не обладает, даже имея постоянное целенаправленное усиление информационного базиса, что характерно практически для всех процедур системно-аналитических исследований. Непосредственное использование аналогов критериев селекции МГУА в такой процедуре не представляется возможным.

Существенный недостаток представленной традиционной схемы САИ, особенно когда это касается сложных проектов, – медленная сходимость результата к требуемым его параметрам. Это приводит к большим затратам времени и средств, а в условиях ограниченности таких ресурсов может и вовсе не привести к желаемым результатам. С этим на практике сталкиваются многие профессиональные коллективы аналитиков.

Используя один из главных принципов методологии МГУА – принцип внешнего дополнения [6, 7], решение этой проблемы может быть достаточно простым и эффективным: в САИ должно быть задействовано *не менее двух экспертно-аналитических групп*, каждая из которых должна работать независимо по «одно-

групповой» схеме. Это равнозначно тому, как в МГУА происходит разделение исходной таблицы статистических данных на две части: A и B , информация каждой из которых служит внешним дополнением друг друга, т.е. обучение модели происходит по индуктивной процедуре на одной части таблицы данных, а ее качество оценивается на другой части и наоборот.

На рис. 2 представлена структура индуктивной процедуры САИ, в которой применена схема многоэтапных алгоритмов МГУА с использованием критериев селекции, обладающих свойствами внешнего дополнения. В этой схеме работают параллельно две группы аналитиков, т.е. реализуется индуктивная «двухгрупповая» процедура системно-аналитического исследования. Эта процедура в крупных чертах была представлена в [8], а из впервые представленной ниже схемы (рис. 2) очевидна ее аналогия с многоэтапным алгоритмом МГУА.

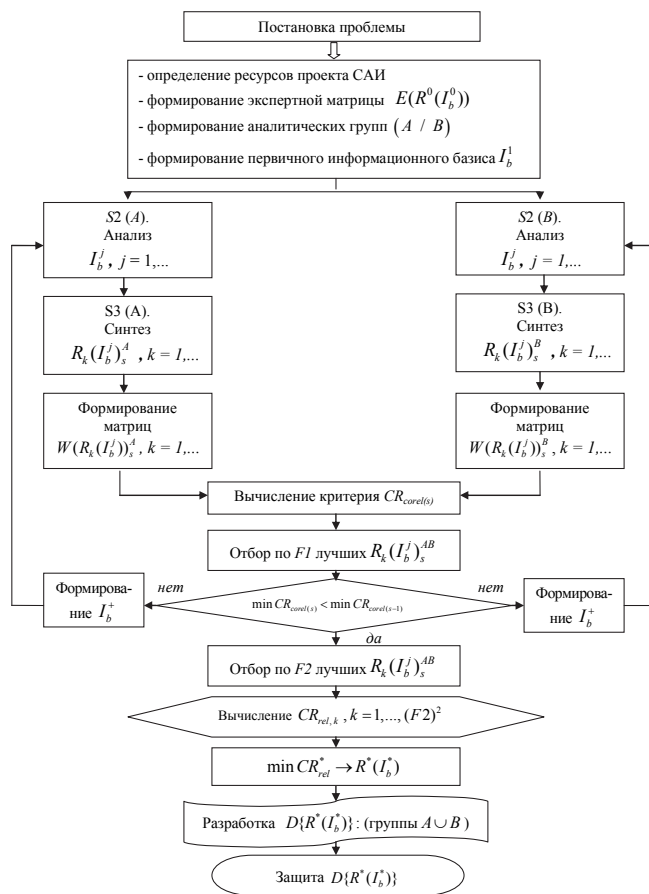


Рис. 2. Общая схема индуктивной процедуры САИ по методологии МГУА

Согласно представленной на рис. 2 процедуре, кроме критерия (4) автора статьи интересует еще некоторый критерий сравнения качества результатов $R_k(I_b^j)_s^A$ и $R_k(I_b^j)_s^B$, полученных двумя независимыми группами (A/B) аналитиков.

Предположим, что алгоритмы формирования экспертной матрицы «целевого» результата $E(R^0(I_b^0))$ и матриц промежуточных частных результатов $W(R_k(I_b^j)_s^A)$, $W(R_k(I_b^j)_s^B)$, $k = 1, \dots$ известны и с их помощью сформирована единственная матрица $E = (e_{ij})$ – для «целевого» результата. Очевидно, что, где это необходимо, могут быть также построены матрицы оценок для частных результатов $W_k^A = (w_{ij}^A)$ и $W_k^B = (w_{ij}^B)$, $k = 1, \dots$. Матрицы $W(R_k(I_b^j)_s^A)$ и $W(R_k(I_b^j)_s^B)$ для процедуры САИ, сконструированной по методологии МГУА, показаны в соответствующих блоках (рис. 2).

Практически во всех или, по крайней мере, в преобладающем большинстве многоэтапных алгоритмов МГУА используется критерий минимума смещения, один из простых и понятных видов которого такой [7]:

$$AR_{bs}^2 = \sum_1^{N_B} (\hat{q}_A - \hat{q}_B)^2 / (\sum_1^N q_O)^2 \rightarrow \min, \quad (6)$$

где \hat{q}_A, \hat{q}_B – выходы моделей одинаковой структуры, коэффициенты которых вычислены на частях A и B заданной выборки статистических данных соответственно, q_O – фактические значения зависимой переменной (выхода), взятые из этой же выборки. Скрупулезные исследования этого критерия приведены в [9].

Для применения критерия (6) в САИ необходимо использовать главное его требование, чтобы модели одинаковой структуры, коэффициенты которых вычислены на частях A и B выборки данных, давали максимально близкие выходы (имели минимальное расхождение по q – отсюда и его название). Используем это замечательное свойство критерия (6) и введем понятие *коррелевантности* в индуктивных САИ.

Аналогично (2), обозначим через Δ_{corel}^2 матрицу

$$\Delta_{corel}^2 = \begin{pmatrix} \delta_{11}^2 & \dots & \delta_{1n}^2 \\ \vdots & \ddots & \vdots \\ \delta_{1m}^2 & \dots & \delta_{mn}^2 \end{pmatrix} = (\delta_{ij}^2), i=1, \dots, n, j=1, \dots, m, \quad (7)$$

элементы δ_{ij}^2 которой равны квадратам разностей элементов матриц $W(R_k(I_b^j))^A$ и $W(R_k(I_b^j))^B$ (обе матрицы также имеют одинаковые размерности $n \times m$, а элементы δ_{ij} относятся к одним и тем же оценкам результата).

Определение 4. Под *корелевантностью* частных результатов (минимума смещения) в индуктивных процедурах САИ понимаем меру вида:

$$CR_{corel} = \|W(R_k(I_b^j))^A - W(R_k(I_b^j))^B\|, \quad (8)$$

а критерием корелевантности в индуктивных процедурах САИ будем называть выражение:

$$CR_{corel} = \left(\sum_{i=1}^m \sum_{j=1}^n (\delta_{ij}^2)_{WW} \right)^{\frac{1}{2}}, \quad (9)$$

где $(\delta_{ij}^2)_{WW}$ – элементы матрицы (7).

Таким образом, если воспользоваться терминологией МГУА, то в приведенной на рис. 2 «двухгрупповой» индуктивной процедуре САИ критерий корелевантности является критерием более низкого уровня и служит для выбора лучших $F1$ частных решений, полученных независимо двумя группами аналитиков A и B , а системный критерий релевантности – критерий верхнего уровня, задача которого заключается в окончательном отборе оптимального результата $R^*(I_b^*)$, который при заданных проектных ресурсах, условиях и требованиях может быть достигнут в проекте системно-аналитического исследования.

Заключение. Рассмотрен один из ключевых вопросов комплексных системно-аналитических исследований – критерии оценки и выбора оптимальных решений. Такие критерии предлагается строить, используя понятие релевант-

ности, которое, как показано, имеет существенные отличия от традиционного его понимания в сфере компилятивного информационного сервиса.

Качество построения и информативность экспертной матрицы, очевидно, занимает одну из центральных позиций при использовании предложенных критериев в развиваемом автором подходе индуктивного проектирования САИ. Поэтому целесообразно было бы сосредоточить внимание дальнейших исследований в данном направлении на таких вопросах как: разработка процедур оценивания уровня компетентности экспертов; выделение однородных групп экспертов или близких оценок для оптимизации структур матриц E и W , а также алгоритмов синтеза таких матриц.

1. Закон України «Про інформацію». – <http://www.rada.gov.ua>
2. Олецкий О.В. Принципи аналізу математичних моделей формування мір подібності та релевантності в рамках тематичного веб-порталу. – www.nbuv.gov.ua/portal/natural/znpipm
3. Осипенко В.В. Принципи методу групового урахування аргументів в контурі системного аналітичного проекту / Індуктивне моделювання складних систем. Вип. 2. – К.: МННЦІТіС НАН України, 2010. – С. 120–129.
4. Батыршин И.З. Основные операции нечеткой логики и их обобщения. – Казань: Отечество, 2001. – 102 с.
5. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. – К.: Слово, 2008. – 344 с.
6. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. – Киев: Наук. думка, 1982. – 296 с.
7. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка, 1985. – 216 с.
8. Осипенко В.В. Синтез результатов в системных проектах аналітичних досліджень за принципами Методу групового урахування аргументів / Індуктивні системи прийняття рішень і проблеми обчислювального інтелекту. – Херсон: ХНТУ, 2010. – С. 117–121.
9. Степашко В.С. Теоретичні аспекти МГУА як методу індуктивного моделювання // УСИМ. – 2003. – № 2. – С. 31–44.

Поступила 02.02.2010
Тел. для справок: (044) 530-5949, (050) 411-7277 (Киев)
E-mail: vvo7@ukr.net
© В.В. Осипенко, 2012