

АДАПТИВНЫЙ ПОДХОД К ОБРАБОТКЕ ДАННЫХ ЭКСПЕРТНОГО ОЦЕНИВАНИЯ ПРИ РЕШЕНИИ ЗАДАЧ В СФЕРЕ ЗАЩИТЫ ИНФОРМАЦИИ

Александр Архипов, София Архипова

Экспертные процедуры являются эффективным, а в ряде случаев и единственным способом получения информации, необходимой для решения широкого круга задач в сфере защиты информации. В статье анализируются точностные аспекты обработки данных групповой экспертизы, в частности, возможности реализации оптимальной обработки этих данных. Показано, что попытка формулировки и решения задачи оптимизации в максимально полной постановке приводит к возникновению конфликтной ситуации, обусловленной с отсутствием необходимых априорных сведений для корректного решения задачи в этой постановке. Выходом является применение адаптивного подхода к организации обработки данных, суть которого состоит в извлечении недостающей информации непосредственно из обрабатываемой совокупности данных. Полученная дополнительная информация в той или иной степени восполняет дефицит априорных сведений, обеспечивая совершенствование структуры и функций элементов процедуры обработки. Совершенствование процедуры обработки повышает ее эффективность и позволяет еще больше расширить объем получаемых дополнительных сведений, обуславливая реализацию нового шага адаптации. В итоге формируется итеративно-процессная схема адаптивной обработки данных, вводится понятие уровня адаптации, рассматривается его методологическое и операциональное наполнение, модельные представления базовых элементов процедуры адаптации. Описаны возможные уровни адаптации, их содержательный аспект, способы накопления и хранения информации о свойствах экспертных данных и характеристиках экспертов, рассмотрены критерии и методология принятия решений о методах обработки данных и формах их реализации.

Ключевые слова: групповая экспертиза, экспертные данные, адаптация, адаптивный подход, компетентность, метод наименьших квадратов (МНК), метод наименьших модулей (МНМ), принцип максимального правдоподобия (ПМП), вариационно-взвешенный МНК.

Введение

Для широкого круга недостаточно формализованных проблем в различных областях человеческой деятельности экспертные процедуры являются эффективным, а в ряде случаев и единственным средством их решения. Оперативность, доступность, универсальность и простота получения релевантных сведений с помощью экспертных методов и процедур обеспечивает их применение в самых разных областях человеческой деятельности, в том числе и такой специфической, как сфера защиты информации (ЗИ) [1, 2, 3].

Как правило, для ЗИ характерны:

а) традиционные экспертизы, регламентированные специальными нормативными или руководящими документами, проводимые для определения степени важности и ценности сведений с целью выявления информации с ограниченным доступом и ее последующей классификации [3];

б) экспертизы, осуществляемые в рамках анализа и оценивания информационных рисков и их компонентов, в ходе проектирования и создания систем ЗИ, общей оценки уровня информационной безопасности организаций и т. п. [1];

в) прочие экспертизы, проводимые для решения различные задач, актуальных для обеспечения эффективной ЗИ [2].

Среди экспертных процедур, применяемых для решения самых разных задач, наиболее популярна коллективная (групповая) экспертиза, позволяющая обеспечить высокую степень точности и объективности конечных результатов. При этом распространен вид групповой экспертизы, для которой общая совокупность экспертных оценок формируется из независимых индивидуальных оценок, полученных от каждого эксперта в отдельности, что обуславливает относительную простоту и прозрачность обработки совокупности данных, генерируемых всей экспертной группой (ЭГ) [4].

Для экспертиз, проводимых в сфере ЗИ характерна сессионность: в ходе работы одной и той же ЭГ из N экспертов этим составом экспертов осуществляется серия из M групповых экспертиз - экспертная сессия. Задачами, решаемыми в ходе экспертной сессии, могут быть: составление перечня возможных угроз, оценка вероятностей реализации отдельных угроз и уровней обусловленных ими потерь, определение количественных показателей надежности и эффективности защитных мероприятий и т. п.

Полагая, что результаты экспертиз имеют количественную форму представления, полученные в ходе сессии данные сводятся в общую матрицу вида:

$$Z = [z_{ij}] = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1N} \\ z_{21} & z_{22} & \dots & z_{2N} \\ \dots & \dots & \dots & \dots \\ z_{M1} & z_{M2} & \dots & z_{MN} \end{bmatrix} = [Z_1, Z_2, \dots, Z_N]. \quad (1)$$

Результаты индивидуальных экспертиз, проведенных j -ым экспертом в ходе сессии, образуют случайный вектор $Z_j = [z_{1j}, z_{2j}, \dots, z_{Mj}]^T$, а данные групповой экспертизы i -ого объекта соответствуют i -ой строке $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$ матрицы (1), каждый элемент которой содержит неизвестное «истинное» значение x_{i0} , общее для всех экспертных оценок z_{ij} и погрешность ε_{ij} , индивидуальную для каждого конкретного эксперта, величина которой зависит от уровня его компетентности:

$$z_{ij} = x_{i0} + \varepsilon_{ij}, \quad i = \overline{1, M}, \quad j = \overline{1, N}. \quad (2)$$

Очевидно, что цель обработки данных каждой отдельной групповой экспертизы – определение неизвестного значения x_{i0} . Для осуществления этой цели необходимо выбрать метод обработки данных экспертизы. Процедура выбора метода включает в себя несколько этапов.

Первый – составление (фиксация) описательной модели формирования данных экспертного опроса, включающей сведения о профессионально-компетентностных характеристиках экспертов, предположения о механизмах генерации ошибок экспертов и свойствах этих ошибок.

Второй этап – разработка (на основе сведений первого этапа) формально-теоретической модели ошибок данных групповой экспертизы.

Третий этап – выбор, с учетом характеристик ошибок и вида их модельного представления методологии обработки данных.

Четвертый этап – выбор (синтез, разработка, формирование) метода обработки данных групповой экспертизы. Выбор метода обычно осуществляется исходя из формально-теоретических моделей и методологических положений, описывающих механизм формирования данных групповой экспертизы (в частности, ошибок этих данных).

К сожалению, на сегодняшний день в практике обработки экспертных данных сложилась и закрепилась следующая негативная тенденция: специалист, которому предстоит использовать экспертные данные, обычно сам решает вопросы

обоснования выбора модели и методологии, действуя при этом весьма субъективно, фактически подстраивая свои выводы под уже намеченный им метод обработки. При этом в приоритете оказываются методы, которые, во-первых, достаточно прозрачны и просты в понимании для обработчика, и, во-вторых, не требуют использования сложных технологий обработки экспертных данных, привлечения к процедуре обработки дополнительной информации, отсутствующей у обработчика в прямом доступе. Другими словами, сложилась ситуация, в которой точностные аспекты обработки экспертных данных, являясь формально приоритетными, фактически уступают место устремлениям обработчика к упрощению процедуры обработки, минимизации интеллектуальных затрат для ее оптимизации.

Постановка задачи

Цель данной статьи – рассмотрение точностных аспектов (качества) обработки данных групповой экспертизы, в частности, возможности оптимизации качества обработки в зависимости от способа описания и характера сведений о механизме возникновения ошибок экспертов, их компетентности и применяемых методах обработки данных. Учитывая обычно существующую у обработчика неполноту перечисленных выше сведений, нередко латентный характер их представления, перспективы реальной оптимизации обработки данных групповой экспертизы в практической ситуации весьма ограничены.

Выходом из создавшейся ситуации является применение адаптивного подхода к обработке данных, суть которого состоит в извлечении из обрабатываемой совокупности данных дополнительной информации, в той или иной степени восполняющей дефицит недостающих априорных сведений, что приводит к повышению эффективности выполняемой обработки за счет совершенствования структуры и функций элементов исходной процедуры обработки. Суть адаптивного подхода – целенаправленное изменение (адаптация) процедуры обработки данных, в частности, ее основного элемента – метода, направленное на повышение уровня оптимальности обработки. В свою очередь совершенствование процедуры обработки позволяет еще больше расширить объем получаемых дополнительных сведений, обуславливая возможность реализации нового шага адаптации процедуры обработки и т. д., формируя в итоге итеративно-процессную схему обработки данных.

В статье предпринята попытка анализа механизма функционирования этой схемы при реализации адаптивного подхода к обработки экспертных данных. В частности, вводится понятие уровня адаптации, рассматривается его методологическое и операциональное наполнение, модельные представления базовых элементов процедуры адаптации, ее валидность и эффективность. Описаны возможные уровни адаптации, их содержательный аспект, формы накопления и хранения информации о свойствах экспертных данных, характеристиках экспертов, методы и способы обработки данных.

Основная часть исследования

Нулевой уровень адаптации. Вводимая для этого уровня описательная модель ЭГ носит идеализированный характер: предполагается равный и высокий уровень компетентности состава экспертов, обеспечивающий несмещенность индивидуальных экспертных оценок с незначительными независимыми случайными ошибками, имеющими произвольное распределение с примерно одинаковыми дисперсиями.

Принятие подобной идеализированной модели позволяет утверждать, что выборка экспертных данных, полученных в ходе групповой экспертизы i -ого объекта и представленных элементами строки z_i матрицы Z , распределена нормально (т.е. по закону Гаусса). В этом случае, исходя из принципа максимального правдоподобия (ПМП), теоретически наилучшей (наиболее достоверной оценкой) неизвестного «истинного» значения x_{i0} будет среднее

$$\bar{z}_i = \frac{1}{N} \sum_{j=1}^N z_{ij}, \quad (3)$$

называемое также оценкой метода наименьших квадратов (МНК-оценкой). Она является несмещенной, эффективной, нормально распределенной и гарантирует минимум суммы квадратов невязок

$$Q_{\text{МНК}} = \sum_{j=1}^N (z_{ij} - \bar{z}_i)^2. \quad (4)$$

Таким образом, на нулевом уровне адаптации обработки экспертных данных получению итогового результата предшествует принятие следующего набора утверждений:

– содержательно-описательная модель ЭГ определяется в виде совокупности N экспертов с равновысоким уровнем компетентности;

– подлежащие обработке данные получены в ходе проведения групповой экспертизы i -ого объекта и представляют собой выборочную совокупность, составленную из N независимых элементов, представленных i -ой строкой $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$ матрицы Z ;

– модельному распределению ошибок экспертов соответствует нормальное распределение (закон Гаусса), что фактически является следствием, выводимым из двух предыдущих допущений;

– учитывая использование статистических моделей для описания ошибок данных, для оптимизация решения задачи выбора метода обработки экспертных данных используется методология ПМП, что, с учетом изложенных выше утверждений, приводит к МНК, причем МНК-оценкой неизвестного значения x_{i0} в данном случае оказывается средневыборочное значения \bar{z}_i элементов i -ой строки матрицы Z .

Фактически наименование «нулевой уровень адаптации» отражает существование характерной для начального этапа обработки экспертных данных конфликтной ситуации – когнитивного противоречия, обусловленного с одной стороны необходимостью максимально полной и общей постановки задачи оптимизации обработки данных, а с другой – отсутствием на практике необходимого объема информации, обеспечивающего целенаправленную адаптацию процедуры обработки для реализации корректного решения поставленной задачи.

Простой и быстрый способ разрешения описанного конфликта состоит в субъективном домысливании отсутствующей информации, позволяющем конкретизировать и сузить неопределенность в исходной постановке задачи и искусственно сформировать информационный базис, обеспечивающий доказуемость заранее предусмотренного решения. Осуществляется подбор на первый взгляд правдоподобных или «нейтральных» допущений, ведущих к достаточно произвольному описанию ошибок данных, к подгонке (или прямой фальсификации) формально-теоретической модели ошибок и, как следствие, – к имитации доказуемости оптимальности заранее выбранного метода обработки.

Один из наиболее распространенных вариантов реализации этого способа – легендирование нормальности ошибок в экспертных данных, заве-

домо ведущее к простейшему, но во многих случаях не лучшему решению задачи оптимизации – вычислению выборочного среднего.

Первый уровень адаптации. Опыт практической работы свидетельствует о том, что представленная выше описательная модель ЭГ фактически не реализуема, так как крайне сложно подобрать группу высококвалифицированных экспертов, близких друг другу по уровню своей компетентности. Группы, включающие экспертов разного уровня, допускают несколько модельных описаний.

Во-первых, это так называемая смесевая модель ЭГ, предполагающая наличие двух групп экспертов: кластера высококвалифицированных специалистов и группы «аномальных» экспертов (АЭ), экспертные оценки которых «выпадают» из общей совокупности данных. Соответственно экспертные данные, полученные от АЭ, могут искажать модель распределения всей остальной совокупности выборочных данных. Идентификация АЭ и устранение из общей матрицы Z всех принадлежащих им данных переводит смесевую модель в описательную модель нулевого уровня адаптации. Однако процедура идентификации АЭ, обычно опирающаяся на применение субъективных решений и эвристических методов анализа, является достаточно сложной, трудоемкой и противоречивой, из-за чего редко применяется на практике.

Кроме того, в условиях постоянного дефицита экспертов, характерного для сферы ЗИ, сокращение состава ЭГ, уже прошедшей экспертную сессию, и, соответственно, выведение из обработки части экспертных данных (полученных от АЭ) крайне нежелательно.

Более перспективной является обобщенная смесевая модель (ОСМ), фактически легитимизирующая реальность появления в составе ЭГ экспертов с разным уровнем компетентности и, как следствие, наличие ощутимых различий в уровнях $\sigma_{\varepsilon_j}^2$ дисперсий ошибок экспертов (гетероскедастичность ошибок). В общем случае ОСМ предполагает существование множественности возможных индивидуальных (частных) моделей распределения ошибок экспертов, что обуславливает проблематичность построения некоторого универсального статистического подхода к обработке экспертных данных. Однако, оставаясь в рамках весьма распространенного предположения о нормальности распределения ошибок каждого эксперта в отдельности, гетероскедастичность

данных в выборке-строке $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$ можно трактовать как следствие случайного варьирования дисперсии Σ_{ε}^2 в законе нормального распределения вида

$$f(\varepsilon) = \frac{1}{\Sigma_{\varepsilon} \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\Sigma_{\varepsilon}^2}}, \quad (5)$$

в который среднеквадратическое отклонение СКО Σ_{ε} входит в качестве случайного параметра. Реализация значений этого параметра, определяющего уровень компетентности конкретного эксперта, происходит в момент включения соответствующего эксперта в ЭГ. В результате для ЭГ получаем вектор-строку реализаций частных значений СКО $[\sigma_{\varepsilon 1}, \sigma_{\varepsilon 2}, \dots, \sigma_{\varepsilon N}]$, характеризующих компетентностный состав ЭГ. Учитывая, что СКО – параметр масштаба, случайную погрешность E произвольной групповой экспертизы можно промоделировать [5] произведением двух взаимно независимых случайных величин Σ_{ε} и E_S

$$E = \Sigma_{\varepsilon} E_S, \quad (6)$$

где E_S - случайная величина, имеющая стандартное нормальное распределение (с нулевым математическим ожиданием и единичным СКО)

$$f_S(\varepsilon_S) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\varepsilon_S^2}{2}}, \quad (7)$$

значения которой масштабируются случайной величиной Σ_{ε} . Для распределения плотности вероятности случайной величины E справедливо соотношение:

$$f(\varepsilon) = \int_{-\infty}^{\infty} f(\sigma_{\varepsilon}) f_S\left(\frac{\varepsilon}{\sigma_{\varepsilon}}\right) \frac{d\sigma_{\varepsilon}}{|\sigma_{\varepsilon}|}. \quad (8)$$

Задание плотности вероятности $f(\sigma_{\varepsilon})$ осуществляется исходя из некоторых правдоподобных (логичных, состоятельных, рациональных) допущений о компетентностных характеристиках экспертов, включенных в состав ЭГ. В частности, несостоятельным представляется предположение о существовании ЭГ из абсолютно компетентных экспертов, суждения (оценки) которых не содержат ошибок (т.е. СКО их ошибок равно 0). Точно также маловероятно (при более-менее ответственном подборе состава экспертов) преобладание в составе ЭГ аномальных экспертов, для которых характерен очень высокий уровень ошибок. Зато вполне правдоподобно наличие в состав ЭГ ряда

достаточно квалифицированных специалистов, чьи СКО локализуются в области значений справа от точки $\sigma_\varepsilon = 0$ в виде некоторого одноэкстремального фрагмента плотности вероятности $f(\sigma_\varepsilon)$. Значение и положение этого экстремума зависят от среднего уровня компетентности экспертов, входящих в ЭГ: снижение этого уровня смещает экстремум вправо, расширяя область локализации возможных значений СКО в сторону больших отклонений. В целом для описания уровня разброса значений СКО ошибок экспертов, включенных в состав ЭГ, достаточно адекватным представляется его аппроксимация распределением Рэлея

$$f(\sigma_\varepsilon) = \frac{\sigma_\varepsilon}{\Delta^2} e^{-\frac{\sigma_\varepsilon^2}{2\Delta^2}}, \quad (9)$$

зависящего от одного параметра Δ , представляющего моду этого распределения. При малых значениях Δ распределение среднеквадратических отклонений σ_ε экспертов, входящих в состав ЭГ, в основном компактно локализуется вдоль оси абсцисс справа от начала координат (рис. 1), что характерно для ЭГ, состоящей из квалифицированных высококомпетентных экспертов. С ростом значений Δ общий уровень компетентности группы снижается за счет появления экспертов, допускающих высокий уровень разброса ошибок, а при $\Delta \geq 2$ в составе группы с достаточно большой вероятностью могут оказаться АЭ.

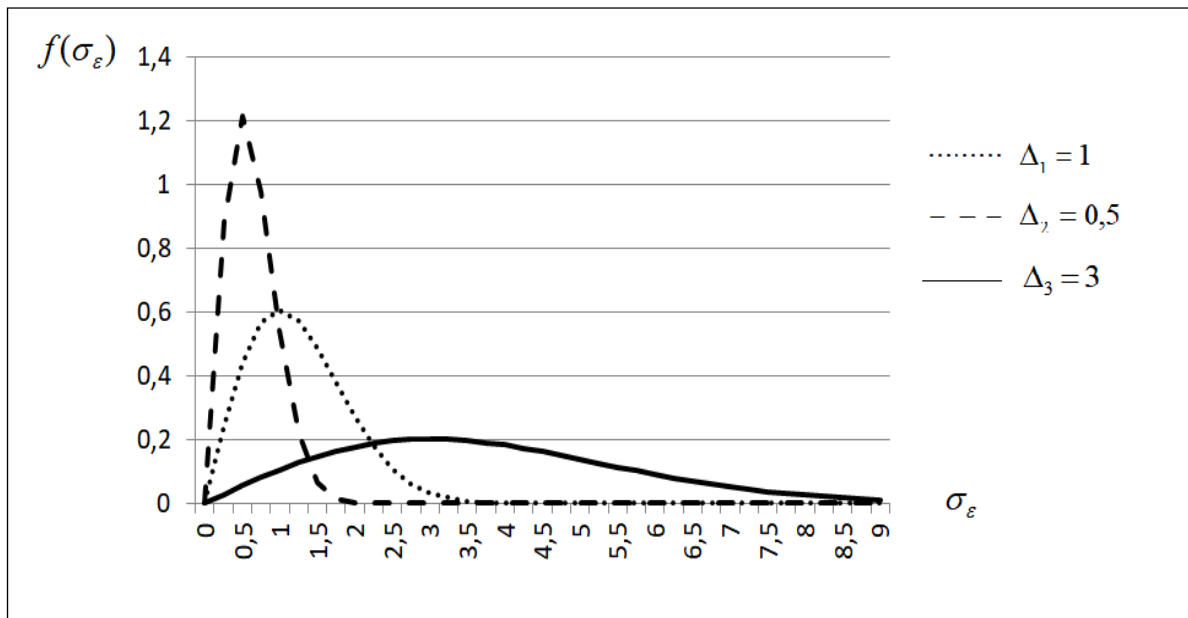


Рис. 1. Аппроксимация распределение $f(\sigma_\varepsilon)$ плотности вероятности СКО ошибок экспертов законом Рэлея для разных значений параметра Δ

Подставляя в выражение (8) соотношения для плотностей вероятностей (7), (9), получаем:

$$f(\varepsilon) = \frac{1}{\Delta^2 \sqrt{2\pi}} \int_0^\infty e^{-\frac{\sigma_\varepsilon^2}{2\Delta^2}} e^{-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}} d\sigma_\varepsilon. \quad (10)$$

Интеграл (10) является частным случаем табличного интеграла [6, С. 356, формула 3.478.4], нахождение которого требует вычисления цилиндрических функций. Опуская промежуточные вычисления, приводим окончательный результат [5]:

$$f(\varepsilon) = \frac{1}{2\Delta} e^{-\frac{|\varepsilon|}{\Delta}}, \quad (11)$$

что соответствует выражению для плотности случайной величины, распределенной по закону Лапласа (рис. 2). В отличие от выражения (5), формула (12) уже не содержит случайного параметра.

Таким образом, результаты групповой экспертизы, выполненной группой экспертов с разным уровнем компетентности, представленные вектором-строкой $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$, можно рассматривать как выборку данных, ошибки наблюдения которых являются реализациями случайной величины, описываемой смесевым распределением равновероятных нормально распределенных эле-

ментов, СКО которых изменяется случайным образом в соответствии с законом Рэлея. В конечном итоге эти ошибки оказываются распределенными по закону Лапласа (11), т.е. отход от идеализированного представления компетентностных характеристик ЭГ и попытка адаптации к реальности приводит к изменению статистической модели ошибок экспертов. При этом, продолжая анализ данных в рамках методологии ПМП, приходим к

новому результату: для лапласового распределения ошибок экспертных данных наилучшей выборочной оценкой значения x_{i0} является медиана z_{imed} , называемая оценкой метода наименьших модулей (МНМ-оценкой), т.к. минимизирует значение статистики вида

$$Q_{МНМ} = \sum_{j=1}^N |z_{ij} - z_{imed}|. \quad (12)$$

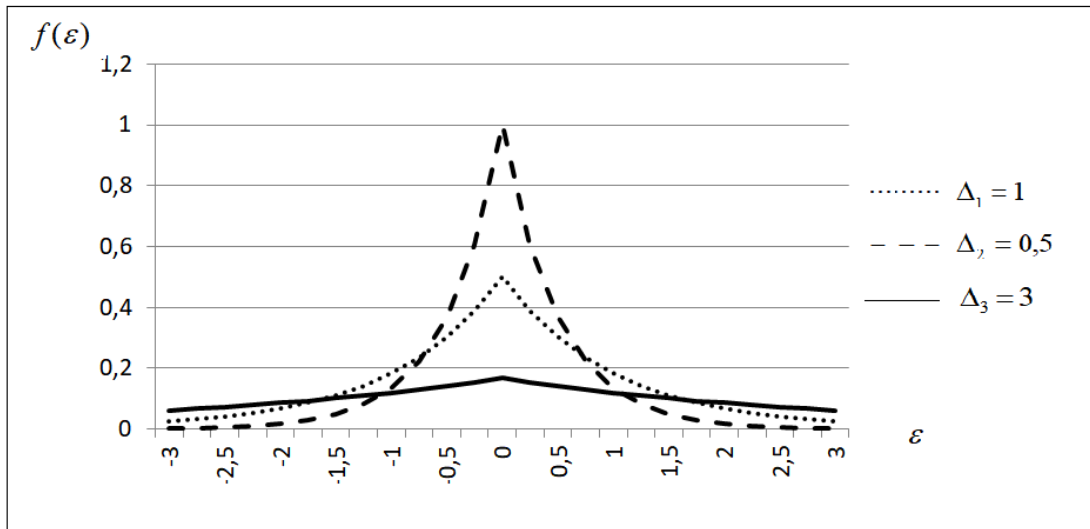


Рис. 2. Распределение $f(\varepsilon)$ ошибок экспертов (закон Лапласа) для разных значений параметра Δ

В виду невозможности непосредственной реализации аналитической процедуры вычисления МНМ-оценок, в ряде случаев в качестве оценки z_{imed} выступает выборочная медиана – средний элемент ранжированной совокупности значений экспертных данных (при нечетном числе N экспертов) или полусумма пары средних элементов (при четном N). В литературе по анализу и обработке данных эта оценка получила наименование «робастной» из-за ее низкой чувствительности к возможным «засорениям» исходной выборки, например, наличие в ней аномальных данных. К сожалению, при малых значениях N применение выборочной медианы может приводить к значительным погрешностям, более точные результаты получаются при аналитическом нахождении МНМ-оценок с помощью итеративного вариационно-взвешенного МНК, при описании которого, чтобы избежать использования двойной индексации, введем замену переменных: $[z_{i1}, z_{i2}, \dots, z_{iN}] = [z_1, z_2, \dots, z_N]$.

Итеративный вариационно-взвешенный МНК, нахождение оценки x_0 [7]:

– **первый шаг:**

$$x_0^{(1)} = \frac{1}{N} \sum_{j=1}^N z_j, \quad \omega_j^{(1)} = \frac{1}{|z_j - x_0^{(1)}|}, \quad (13)$$

где $x_j, j = \overline{1, N}$ – данные, полученные от экспертов в ходе групповой экспертизы;

– **второй шаг:**

$$x_0^{(2)} = \frac{1}{\sum_j \omega_j^{(1)}} \sum_j z_j \omega_j^{(1)}, \quad \omega_j^{(2)} = \frac{1}{|z_j - x_0^{(2)}|}, \quad (14)$$

– **k-ый шаг:**

$$x_0^{(k)} = \frac{1}{\sum_j \omega_j^{(k-1)}} \sum_j z_j \omega_j^{(k-1)}, \quad \omega_j^{(k)} = \frac{1}{|z_j - x_0^{(k)}|}. \quad (15)$$

Если окажется, что очередные результаты, полученные на текущем k -ом шаге, практически совпадают со значениями, рассчитанными на предыдущем шаге, т.е. $x_0^{(k-1)} \approx x_0^{(k)}, \omega_j^{(k-1)} = \omega_j^{(k)}, j = \overline{1, N}$, итеративный цикл прекращается, и значение $x_0^{(k)}$ принимается в качестве искомой МНМ-оценки z_{imed} .

Таким образом, первый уровень адаптации характеризуется отходом от идеализированного

описания состава ЭГ, попыткой учета существования реальных различий в компетентностях экспертов, приводящей в итоге к трансформации модельного закона распределения элементов выборки-строки $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$ из нормального в закон Лапласа, что, в соответствии с методологией ПМП, обуславливает изменения применяемого метода оценивания: переход от МНК- к МНМ-оценкам. К сожалению, обоснование закона Лапласа опирается на субъективное решение, связанное с предположением о возможности описания случайного разброса СКО ошибок экспертов законом Рэлея.

Второй уровень адаптации. При реализации обработки данных с нулевым и первым уровнем адаптации для выбора метода обработки использовалась только информация общего характера о предположительных особенностях механизма формирования случайных ошибок в данных групповой экспертизы. Однако проведение экспертной сессии и, как результат, наличие полного объема сведений о проводимой в ее рамках серии групповых экспертиз, представленных матрицей Z , позволяет частично отказаться от использования вероятностно-статистического подхода к описанию и обработке экспертных данных. Предлагается персонифицировать подход к обработке данных групповой экспертизы, применив более простой и прозрачный метод взвешенного среднего, где веса зависят от степени доверия к информации, получаемой от каждого конкретного эксперта. Основная проблема, возникающая в этом случае, – объективное назначение весов.

При определении весов исходим из того, что каждый вектор-столбец $Z_j = [z_{1j}, z_{2j}, \dots, z_{mj}]^T$, $j = \overline{1, N}$ матрицы Z , составленный из результатов индивидуальных экспертиз, проведенных j -ым экспертом, содержит информацию двух видов: являющуюся целью экспертизы «правильную» детерминированную информацию x_{j0} о каждом j -ом экспертируемом объекте, не известную, но общую для всех экспертов, и информацию об индивидуальных случайных ошибках эксперта, характеризующих уровень его индивидуальной компетентности. При отсутствии ошибок экспертов результаты индивидуальных экспертиз должны совпадать и могли бы быть представлены одной точкой $X_0 = [x_{10}, x_{20}, \dots, x_{M0}]^T$ в M -мерном пространстве результатов экспертной сессии. Однако наличие ошибок экспертов приводит к расщеплению этой

точки в облако (кластер), содержащее N точек-образов отдельных экспертов. При адекватном подборе состава ЭГ плотность облака неоднородна и максимальна в окрестности, прилегающей к центру кластера Z_0 . Расстояние r_j от точки $X_0 = [x_{10}, x_{20}, \dots, x_{M0}]^T$ до точки-образа j -ого эксперта характеризует уровень компетентности этого эксперта и используется при расчете показателя компетентности [8]:

$$c_j = f(r_j) = \frac{1}{(1-B)e^{b \cdot r_j} + B}. \quad (16)$$

Следует учесть, что количественные оценки, получаемые в различных групповых экспертизах, могут отличаться единицами измерения, масштабом, шкалой. Поэтому прежде чем приступить к вычислению расстояний r_j , $j = \overline{1, N}$, данные матрицы Z (формула (2)) следует пересчитать к некоторой унифицированной форме представления с единообразной шкалой, частным примером которой является балльная шкала. В качестве приближенного значения координат $X_0 = [z_{10}, z_{20}, \dots, z_{M0}]^T$ можно использовать оценки \bar{z}_i , z_{imed} , $i = \overline{1, M}$ центра кластера Z_0 .

Формула (16) определяет структуру шкального преобразования $c_j = f(r_j)$, возможные (рекомендованные [8]) значения параметров: $b_0 = 15$, $B = 0,967$. Оценки расстояния r_j теоретически могут лежать в диапазоне от 0 до значений, близких к ∞ , а соответствующие им значения показателя компетентности c_j изменяются в диапазоне от 1 до 0 (полная, абсолютная некомпетентность). Знание элементов вектора компетентностей экспертов $C = [c_1, c_2, \dots, c_N]$ позволяет рассчитать результирующие средневзвешенные оценки для каждого i -ого объекта групповой экспертизы:

$$\tilde{z}_i = \frac{\sum_{j=1}^N c_j z_{ij}}{\sum_{j=1}^N c_j} = \sum_{j=1}^N w_j z_{ij}, \quad i = \overline{1, M}, \quad (17)$$

где $w_j = c_j / \sum_{j=1}^N c_j$, $j = \overline{1, N}$ – весовые коэффициенты, учитывающие соотношение уровней компетентности экспертов. При «идеальном» подборе экспертов, входящих в состав ЭГ, т.е. для случая $c_j = 1$, $j = \overline{1, N}$, все веса оказываются единичными: $w_j = 1$, оценки среднего и взвешенного среднего совпадают.

Таким образом, второй уровень адаптации обеспечивает достаточно высокую конкретизацию сведений об экспертах, что позволяет исключить применение теоретико-вероятностных моделей как для описания характеристик ЭГ, так и для описания свойств исходных данных. Соответственно вместо методологии ПМП используется понятие ценности (полезности, важности) сведений, получаемых от экспертов, базирующееся на оценках их компетентности, пересчитываемых в систему персонифицированных весовых коэффициентов w_j , $j = \overline{1, N}$, позволяя получить простые и наглядные средневзвешенные результирующие оценки групповой экспертизы для каждого из экспертируемых объектов.

Третий уровень адаптации. Для этого уровня адаптации необходимо наличие точных значений параметров (показателей) оценивавшихся на предшествующих сеансах экспертной сессии. Типичный пример – экспертные прогнозы вероятностей реализации угроз, количественных оценок потерь и т.п., при условии, что эти прогнозы носят периодический или квазипериодический характер и к моменту проведения сеанса очередной групповой $M + 1$ экспертизы уже известны точные значения параметров (показателей), прогнозовавшихся на предшествующем M -ом сеансе экспертной сессии, т.е. в общем случае известен вектор $X_0 = [x_{10}, x_{20}, \dots, x_{M0}]^T$ точных апостериорных значений прогнозируемого параметра.

Согласно методологии ПМП, МНК-оценка представляет собой оптимальную оценку неизвестного «истинное» значение x_{i0} при нормальной (гауссовой) модели распределения ошибок экспертов в элементах выборки-строки $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$, в случае распределения этих ошибок по закону Лапласа оптимальной становится МНМ-оценка. Однако распределение ошибок в данных реальной групповой экспертизы в той или иной мере всегда отличается от модельных распределений, поэтому оценки \bar{z}_i , z_{imed} являются квазиоптимальными, причем оценить степень потери оптимальности не представляется возможным. Тем не менее, основываясь на материалах, представленных в [9], на базе полученных квазиоптимальных оценок можно построить так называемую групповую (комплексную) оценку z_{igr} , дисперсия которой в наихудшем случае не превысит меньшей из дисперсий исходных оце-

нок. Информация, необходимая для формирования оценки z_{igr} , задается ковариационной матрицей ошибок методов оценивания (т.е. ошибок МНК-оценок \bar{z}_i и МНМ-оценок z_{imed} , $i = \overline{1, M}$) вида

$$COV = \begin{bmatrix} D\{\bar{z}\} & cov \\ cov & D\{z_{med}\} \end{bmatrix}, \quad (18)$$

где $D\{\dots\}$ и cov - дисперсии и ковариация ошибок, рассчитываемые по формулам:

$$D\{\bar{z}\} = \frac{1}{M} \sum_{i=1}^M (\bar{z}_i - x_{i0})^2, \quad (19)$$

$$D\{z_{med}\} = \frac{1}{M} \sum_{i=1}^M (z_{imed} - x_{i0})^2, \quad (20)$$

$$cov = \frac{1}{M} \sum_{i=1}^M (\bar{z}_i - x_{i0})(z_{imed} - x_{i0}). \quad (21)$$

По данным ковариационной матрицы COV рассчитываются общие веса для МНК- и МНМ-оценок и значение групповой оценки для каждой отдельной групповой экспертизы:

$$z_{igr} = w_1 \bar{z}_i + w_2 z_{imed}, \quad (22)$$

где

$$w_1 = \frac{D\{z_{med}\} - cov}{D\{\bar{z}\} + D\{z_{med}\} - 2cov}, \quad (23)$$

$$w_2 = \frac{D\{\bar{z}\} - cov}{D\{\bar{z}\} + D\{z_{med}\} - 2cov}. \quad (24)$$

Дисперсия групповой оценки определяется формулой

$$D\{z_{gr}\} = w_1^2 D\{\bar{z}\} + w_2^2 D\{z_{med}\} + 2cov. \quad (25)$$

Нахождение значений элементов матрицы COV оказывается возможным только при наличии совокупности данных, включающей множества пар оценок \bar{z}_i , z_{imed} , $i = \overline{1, M}$, полученных ранее в ходе построчной обработки матрицы Z на двух первых уровнях адаптации, и вектора $X_0 = [x_{10}, x_{20}, \dots, x_{M0}]^T$ точных апостериорных значений прогноза. Совместная обработка этих сведений по формулам (19) – (21) позволяет создать дополнительную информацию для определения элементов матрицы COV , используемых далее для вычисления групповой оценки z_{igr} .

Заметим, что для обработки данных групповой экспертизы, представленных выборкой-строкой $z_i = [z_{i1}, z_{i2}, \dots, z_{iN}]$, могут привлекаться наряду с МНК, МНМ и другие методы. В общем случае результатом применения K различных методов

будет множество оценок $\tilde{z}_{ik}, k = \overline{1, K}$, где k - порядковый номер метода обработки, которым была рассчитана соответствующая k -ая оценка. Для полученного множества значений $\tilde{z}_{ik}, k = \overline{1, K}$ также возможно формирование групповой (комплексной) оценки z_{igr} :

$$z_{igr} = \sum_{k=1}^K w_k \tilde{z}_{ik}. \quad (26)$$

Для нахождения групповой оценки (25) используются следующие рабочие соотношения [9]:

$$W = \frac{COV^{-1} \uparrow}{\uparrow^T COV^{-1} \uparrow}, \quad (27)$$

$$D\{z_{gr}\} = W^T COV W, \quad (28)$$

где $W = [w_1, w_2, \dots, w_K]^T$ - вектор весовых коэффициентов методов обработки, COV - ковариационная матрица ошибок методов обработки (аналогичная матрице (18), симметрическая, размерностью $K \times K$), $\uparrow = [1, 1, \dots, 1]^T$ - вектор-столбец длиной K , все элементы которого - единицы.

Объем сведений, доступных на третьем уровне адаптации, допускает, кроме построения групповых оценок, применение уже рассмотренного на втором уровне адаптации метода взвешенного среднего. При оценивании компетентностей экспертов, благодаря наличию точных апостериорных значений $X_0 = [x_{10}, x_{20}, \dots, x_{M0}]^T$, устраняются возможные ошибки вычисления расстояний $r_j, j = \overline{1, N}$, обусловленные использованием на втором уровне адаптации вместо точных координат точки X_0 в M -мерном пространстве их приближенных оценочных значений, соответствующих координатам Z_0 центра кластера результатов экспертной сессии. Это позволяет повысить точность индивидуальных оценок компетентности экспертов в диапазоне высоких значений компетенции (близких к 1).

Выводы

Адаптивный подход к обработке экспертных данных - целенаправленное формирование (адаптация) процедуры обработки данных, в частности, ее основного элемента - метода, ориентированное на повышение уровня точности обработки. Адаптивный подход не является однократным операциональным актом, так как совершенствование процедуры обработки носит процессный характер, во многом зависящий от форм и способов накопления и хранения информации о свойствах экспертных данных, пониманию особенностей механизма

их формирования, выбора метода обработки, этапа и уровня реализации адаптации, пр.

Для начального этапа обработки экспертных данных характерно наличие конфликтной ситуации - когнитивного противоречия, - обусловленного с одной стороны необходимостью максимально полной и общей постановки задачи обработки, а с другой - отсутствием необходимого объема информации для корректного решения задачи в этой постановке. Возможный способ разрешения этого конфликта состоит в реализации адаптивного подхода к обработке данных, носящего итеративно-процессный характер и операционально представляющего собой последовательность этапов обработки данных, каждый из которых, приводя к решению поставленной задачи с определенной степенью точности, позволяет уточнить, детализировать исходную постановку задачи для оптимизации обработки на последующих этапах. В частности, к элементам детализации может относиться уточнение модели ошибок в экспертных данных, конкретизация уровней компетентности экспертов, другие аспекты построения и организации процедуры обработки данных. Процессный характер учета этих дополнительных сведений ведет к все более высокому уровню адаптации метода обработки экспертных данных к особенностям ошибок данных и в конечном итоге - к получению оптимальной либо почти оптимальной точности обработки.

ЛИТЕРАТУРА

- [1]. Ю. Ткач, С. Казмірчук, Д. Мехед, В. Базилевич, "Застосування методу експертних оцінок до оцінювання інформаційних ризиків вищого навчального закладу", *Захист інформації*, Т. 19, № 2, С. 142-157, 2017.
- [2]. А. Архипов, С. Архипова, С. Носок, "Технологии экспертного оценивания в задачах защиты информации" *Інформаційні технології та комп'ютерна інженерія? Міжнародний науково-технічний журнал* №1, С. 89-94, 2005.
- [3]. О. Корченко, О. Архипов, Ю. Дрейс, *Оцінювання шкоди національній безпеці України у разі витоку державної таємниці: Монографія* Київ. наук.-вид. центр Нац. Академія СБ України, 2014, 332 с.
- [4]. О. Архипов. *Вступ до теорії ризиків: інформаційні ризики: Монографія* Київ. Нац. Академія СБ України, 2015, 248 с.
- [5]. А. Архипов, С. Архипова, "Обобщенная модель случайных погрешностей групповой экспертизы", *Системные технологии, Региональный межвузовский сборник научных работ*, Вып. 4(87), Днепропетровск, С. 3-10, 2013.
- [6]. И. Градштейн, И. Ръжик. *Таблицы интегралов, сумм, рядов и произведений*, Москва. Физматгиз, 1963, 1108 с.

[7]. О. Архипов, О. Чмерук, "Дослідження методів обробки даних багатооб'єктної експертизи" Матеріали XVI Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених. *Теоретичні і прикладні проблеми фізики, математики та інформатики*, т. II (26-27 квіт. 2018 р., м. Київ). Київ. КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», С. 53-57, 2018.

[8]. О. Архипов, С. Архіпова, "Оцінювання якості роботи експертів за даними багатооб'єктної експертизи", *Захист інформації*, Т. 13, №4 (53), С. 45-54, 2011.

[9]. А. Архипов, "Построение оптимальных групповых оценок", *Адаптивные системы автоматического управления*, Киев: Техника, №17, С. 75-78, 1989.

ADAPTIVE APPROACH TO PROCESSING DATA OF EXPERTS EVALUATION WHEN SOLVING PROBLEMS IN THE FIELD OF INFORMATION PROTECTION

Expert procedures are effective, and in some cases the only way to obtain the information necessary to solve a wide range of tasks in the field of information protection. The article analyzes the accuracy aspects of the processing of group examination data, in particular, the possibility of realizing the optimal processing of this data. It is shown that an attempt to formulate and solve the optimization problem in the most complete statement leads to a conflict situation due to the lack of the necessary a priori information for the correct solution of the problem in this statement. The way out is the use of an adaptive approach to organizing data processing, the essence of which is to extract the missing information directly from the processed data set. The additional information received, to one degree or another, makes up for the deficit of a priori information, ensuring the improvement of the structure and functions of the elements of the processing procedure. Improving the processing procedure increases its efficiency and allows you to further expand the amount of additional information received, causing the implementation of a new adaptation step. As a result, an iterative process mode of adaptive data processing is formed, the concept of the adaptation level is introduced, its methodological and operational content, model representations of the basic elements of the adaptation procedure are considered. Possible levels of adaptation, their substantive aspect, methods of accumulating and storing information about the properties of expert data and characteristics of experts are described, the criteria and methodology for making decisions about data processing methods and forms of their implementation are considered.

Keywords: group examination, expert data, adaptation, adaptive approach, competence, least squares method, least modules method, maximum likelihood principle, variationally weighted least squares method.

АДАПТИВНИЙ ПІДХІД ДО ОБРОБКИ ДАНИХ ЕКСПЕРТНОГО ОЦІНЮВАННЯ ПРИ ВИРІШЕННІ ЗАВДАНЬ У СФЕРІ ЗАХИСТУ ІНФОРМАЦІЇ

Експертні процедури є ефективним, а в ряді випадків і єдиним способом отримання інформації, необхідної

для вирішення широкого кола завдань у сфері захисту інформації. У статті аналізуються аспекти точності обробки даних групової експертизи, зокрема, можливості реалізації оптимальної обробки цих даних. Показано, що спроба формулювання і рішення задачі оптимізації в максимально повній постановці призводить до виникнення конфліктної ситуації, обумовленої відсутністю необхідних апріорних відомостей для коректного вирішення завдання в цій постановці. Виходом є застосування адаптивного підходу до організації обробки даних, суть якого полягає в отриманні потрібної додаткової інформації безпосередньо з оброблюваної сукупності даних. Отримана інформація в тій чи іншій мірі поповнює дефіцит апріорних відомостей, забезпечуючи вдосконалення структури і функцій елементів процедури обробки. Удосконалення процедури обробки підвищує її ефективність і дозволяє ще більше розширити обсяг отримуваних додаткових відомостей, обумовлюючи реалізацію нового кроку адаптації. В результаті формується ітеративно-процесна схема адаптивної обробки даних, вводиться поняття рівня адаптації, розглядається його методологічне та операціональне наповнення, модельні представлення базових елементів процедури адаптації. Описано можливі рівні адаптації, їх змістовний аспект, способи накопичення і зберігання інформації про властивості експертних даних і характеристики експертів, розглянуті критерії та методологія прийняття рішень про методи обробки даних і форми їх реалізації.

Ключові слова: групова експертиза, експертні дані, адаптація, адаптивний підхід, компетентність, метод найменших квадратів, метод найменших модулів, принцип максимальної правдоподібності, варіаційно-зважений метод найменших квадратів.

Архипов Олександр Євгенійович, д.т.н., професор, професор кафедри інформаційної безпеки НТУУ «КПІ імені Ігора Сікорського».

E-mail: sonet0515@gmail.com.

Orcid ID: 0000-0001-6832-2223.

Архипов Александр Евгеньевич, д.т.н., професор, професор кафедри інформаційної безпеки НТУУ «КПІ імені Ігора Сікорського».

Arkhipov Olexander, Dr.Sci.Tech., professor at the Department of Information Defense, NTUU «Igor Sikorsky Kyiv Polytechnic Institute».

Архіпова Софія Анатоліївна, к.т.н., доцент, доцент кафедри філософії, НТУУ «КПІ імені Ігора Сікорського».

E-mail: arsofi@ukr.net.

Orcid ID: 0000-0002-3116-0731.

Архіпова Софія Анатоліївна, к.т.н., доцент, доцент кафедри філософії, НТУУ «КПІ імені Ігора Сікорського».

Arkhipova Sofiia, Candidate of Technical Sciences, Associate Professor, Associate Professor of Philosophy, NTUU «Igor Sikorsky Kyiv Polytechnic Institute».