УДК 519.24:004.89

А.Е. Архипов, С.А. Архипова

ОБОБЩЕННАЯ МОДЕЛЬ СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ ГРУППОВОЙ ЭКСПЕРТИЗЫ

Анотація. У статті аналізується розподіл похибки експертно-аналітичних оцінок, що розраховують за даними групової експертизи. У припущенні нормальності розподілу цієї похибки показано, що її середнє квадратичне відхилення (СКВ) виявляється випадковою величиною, значення якої визначаються впливом двох груп факторів: помилок експертів і похибки відображення вихідних даних групової експертизи в результуючу експертно-аналітичну оцінку. При однаковій інтенсивності прояву цих факторів розподіл СКВ описується законом Релея, що спричиняє трансформацію вихідного нормального розподілу похибки в закон Лапласа.

Введение

Последние четверть века в самых разных сферах деятельности наблюдается повышенный интерес к экспертным технологиям обработки информации, в частности, к использованию экспертных оценок в задачах принятия решения, применении экспертных методов для целей диагностики, оптимизации, классификации и прогнозирования [1-4]. Современные методы экспертизы все чаще вместо прямого экспертного оценивания, в ходе которого эксперт непосредственно определяет качественную либо количественную оценку исследуемого качества (свойства, признака) представленного к экспертизе объекта, применяют процедуру экспертно-аналитического продуцирования оценки. Механизм экспертно-аналитического оценивания (ЭАО) базируется на предположении существования отображения $\phi : X \to Y$ исследуемого качества Y объекта экспертизы комплексом других его свойств $X = (X_1, X_2, \ldots, X_n)$, допускающих простое и точное оценивание. Поэтому двухэтапная процедура ЭАО, первый этап которой – экспертное оценивание значений свойств $X_1, X_2, \ldots, X_n$, а второй – получение отображения $\phi : X \to Y$ искомой оценки качества Y, несмотря на очевидное усложнение экспертизы, позволяет надеяться, что в целом ее итоговый результат будет иметь более прозрачный, надежный и объективный характер.

Функции эксперта в процедуре ЭАО обычно сводятся к формированию кортежа $< x_1, x_2, \ldots, x_n >$, содержащего представленные в логической и (или) в количественной форме оценки вектора свойств $X = (X_1, X_2, \ldots, X_n)$. Получаемые оценки часто являются ответами эксперта на вопросы специально составленного вопросника, причем размерность n вектора X может достигать сотен, а порядок следования и характер вопросов могут изменяться в зависимости от особенностей информации, содержащейся в ответах на поставленные ранее вопросы [4]. В связи с этим для проведения опроса эксперта обычно используется ЭВМ, реализующая диалоговую процедуру общения с экспертом. Обработка полученной от эксперта информации, в частности выполнения отображения $\phi : X \to Y$, оказывается обычно весьма сложной и громоздкой задачей, также требующей привлечения ЭВМ.

Поэтому для комплексного обеспечения проведения экспертиз, в первую очередь реализации диалоговой процедуры получения экспертной информации («извлечения» информации «из эксперта») и автоматизации ее последующей обработки используются специализированные информационные системы поддержки принятия решений (ИСППР). При этом область применения ИСППР, масштабы и целевая направленность решаемых проблем могут быть самыми разными, от нахождения точечных экспертных оценок количественных признаков в задачах прогнозирования, оценивания параметров систем и процессов, до обеспечения проведения сложных экспертиз, предназначенных для выбора структуры проектируемых систем, оптимизации их характеристик, ресурсного обеспечения и т.п. [2,3]. Кроме того в круг задач, решаемыми с помощью подобных ИСППР, часто входят задачи, связанные с определением качества (точности, надежности) решений, получаемых при использовании экспертных оценок. Наиболее актуальна эта проблематика в упомянутых выше задачах точечного оценивания количественных характеристик, рассчитываемых по данным, полученным с учетом групповой (коллективной) экспертизы [2,5,6]. Для объективной обработки подобных данных требуется информация об уровнях компетентности экспертов, задействованных в групповой экспертизе. При этом обычно ограничиваются

 ISSN 1562-9945

сведениями о степени интенсивности ошибок, допускаемых каждым экспертом в ходе экспертизы и виде закона распределения ошибок. В связи с индивидуализированным характером подобной информации ее учет возможен лишь на этапе обработки уже полученных экспертных данных. Поэтому в ряде разработок ИСППР, предназначенных для обеспечения экспертиз, предусматривается набор моделей распределений, каждой из которых соответствует своя оценка уровня компетентности эксперта. Предполагается, что выбор модели распределения осуществляется экспертом. Однако вряд ли подобный подход реализуем на практике: во-первых, эксперт-прикладник – это специалист в своей предметной области, возможно не всегда в достаточной области ориентирующийся в теории вероятностей и математической статистике, во-вторых, необходимость задания «персональной» модели ошибок, особенно в условиях проведения групповой экспертизы, является этически некорректным требованием.

Ниже рассматривается попытка построения модели распределения вероятностей погрешностей результатов групповой экспертизы при достаточно общих предположениях о факторах, обуславливающих возникновение этих погрешностей.

Формулировка задачи

Наиболее простой и распространенный способ описания погрешности *E* результата групповой экспертизы – это постулирование нормального закона распределения вероятностей ее значений в предположении несмещенности погрешности (т.е. математическое ожидание $m_\varepsilon = 0$). Предполагается, что нормальность распределения погрешности *E* результирующих оценок групповых экспертиз обусловливается интегральным влиянием множественных ошибок в индивидуальных экспертных оценках. Кроме того считается, что дисперсия погрешности σ_ε^2 является следствием разброса индивидуальных оценок экспертов, разброс этих оценок – единственная причина появления погрешности в результатах групповой экспертизы, из чего следует, что для «идеального» (не ошибающегося) эксперта отображение его вектора оценок X^+ должно привести к абсолютно точному экспертно-аналитическому определению значения исследуемого качества *Y*, т.е. $\phi : X^+ \to Y$.

При этом игнорируется тот факт, что «естественный» механизм отображения ϕ не известен, в ИСППР вместо него работает некоторая аппроксимативная модель α этого отображения, применение которой обуславливает появление погрешности отображения $\alpha : X^{+} \to Z = Y + E_{\alpha}$. Чаще всего погрешность E_{α} является следствием неминуемых потерь информации, имеющих место при коммуникации эксперта и разработчика ИСППР на этапе ее проектирования. По оценкам, приведенным в [6], остаточная информация, получаемая от эксперта в процессе коммуникации, составляет порядка 24% от ее исходного объема. По сведениям [7,8], практически не передаются при коммуникации так называемые знания второго рода – эмпирические правила, интуитивные соображения и факты, которые, как правило, не публикуются, но дают возможность опытному эксперту эффективно принимать решения даже в условиях неполных и противоречивых исходных данных. Очевидно, что погрешность отображения E_{α} – это погрешность, обусловленная безвозвратными потерями информации, для восполнения которой бесполезна любая апостериорная математическая обработка остаточной информации.

Поэтому можно предположить, что возникновение погрешности *E* результирующих оценок групповой экспертизы обусловливается действием двух независимых друг от друга факторов: ошибок, допущенных при реализации отображения α (фактор V_1) и индивидуальных ошибок экспертов (фактор V_2).

Очевидно, что дисперсия σ_{ε}^{2} погрешности *E* при воздействии только одного фактора V_1 оказывается равной σ_1^2, при воздействии только фактора V_2 – σ_2^2. Открытым остается вопрос, как зависит дисперсия σ_{ε}^{2} от совместного воздействия обоих факторов.

Основная часть

Рассмотрим систему случайных величин (Σ_1, Σ_2), реализациями которых являются средние квадратические отклонения (СКО) σ_1, σ_2. Предположим, что каждая из случайных величин, входящих в систему (Σ_1, Σ_2) порождается действием соответствующего фактора V_1 или V_2. При наличии обоих факторов, учитывая их взаимную независимость, и в предположении, что значения σ_1, σ_2 распределены нормально:

$$f_1(\sigma_1) = \frac{1}{\Delta_1\sqrt{2\pi}} e^{-\frac{\sigma_1^2}{\Delta_1^2}}, \qquad f_2(\sigma_2) = \frac{1}{\Delta_2\sqrt{2\pi}} e^{-\frac{\sigma_2^2}{\Delta_2^2}}, \tag{1}$$

вероятность попадания случайной точки (Σ_1, Σ_2) в эллипс рассеяния B_l (с центром в точке (0,0), и с полуосями $x = l\Delta_1$, $y = l\Delta_2$), описываемый уравнением

$$P((\Sigma_1, \Sigma_2) \in B_l) = \int_{B_l}\int 1 / 2\pi\Delta_1\Delta_2 \exp\left\{-\frac{1}{2}\left(\frac{\sigma_1^2}{\Delta_1^2} + \frac{\sigma_2^2}{\Delta_2^2}\right)\right\} d\sigma_1 d\sigma_2 . \tag{2}$$

Переходя к полярным координатам r, γ:

$$\frac{\sigma_1^2}{\Delta_1\sqrt{2}} = r\cos\gamma, \qquad \frac{\sigma_2^2}{\Delta_2\sqrt{2}} = r\sin\gamma , \quad \left(0 < r < \frac{l}{\sqrt{2}} ,\ 0 < \gamma < 2\pi\right), \tag{3}$$

получаем [9 , стр. 238]:

$$P((\Sigma_1, \Sigma_2) \in B_l) = 1 - e^{-\frac{l^2}{2}} . \tag{4}$$

В случае круговой симметрии нормальных распределений (1), т.е. при $\Delta_1 = \Delta_2 = \Delta$, появляется возможность выразить соотношение (4) через параметры исходных распределений (1). В частности, вероятность того, что случайная точка (Σ_1, Σ_2) попадет внутрь (теперь уже круга) B_l с радиусом $\Sigma_\varepsilon = \sqrt{\Sigma_1^2 + \Sigma_2^2}$, определяется формулой (4) при $l = \sigma_\varepsilon / \Delta$, что соответствует функции распределения случайной величины Σ_ε:

$$P((\Sigma_\varepsilon < \sigma_\varepsilon) = F(\sigma_\varepsilon) = 1 - e^{-\frac{\sigma_\varepsilon^2}{2\Delta^2}} . \qquad (\sigma_\varepsilon > 0) . \tag{5}$$

Производная функции распределения (5) дает плотность распределения

$$f(\sigma_\varepsilon) = F'(\sigma_\varepsilon) = \frac{\sigma_\varepsilon}{\Delta^2} e^{-\frac{\sigma_\varepsilon^2}{2\Delta^2}} , \tag{6}$$

называемую законом Рэлея [9,10].

В случае круговой симметрии нормальных распределений (1) получаем возможность простой и наглядной интерпретации распределения погрешности *E* групповой экспертизы. Как уже отмечалось выше, распределение погрешности *E* традиционно предполагается нормальным, однако в выражение плотности распределения

$$f(\varepsilon) = \frac{1}{\Sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{\Sigma_\varepsilon^2}}, \qquad (7)$$

входит случайная дисперсия $\Sigma_\varepsilon{}^2$, распределенная, как это следует из выражений (5), (6), по закону Рэлея, Учитывая, что дисперсия является масштабирующим параметром, погрешность *E* можно рассматривать как случайную величину E_S, имеющую стандартное нормальное распределение (с нулевым математическим ожиданием и единичной дисперсией), значения которой промасштабированы случайной величиной Σ_ε[11], т.е.

$$E = \Sigma_\varepsilon E_S. \qquad (8)$$

Для взаимно независимых случайных величин Σ_ε и E_S справедливо соотношение [9]:

$$f(\varepsilon) = \int_{-\infty}^{\infty} f(\sigma_\varepsilon) f_S\left(\frac{\varepsilon}{\sigma_\varepsilon}\right) \frac{d\sigma_\varepsilon}{|\sigma_\varepsilon|}, \qquad (9)$$

где $f_S(\varepsilon_S) = f_S\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)$ – плотность стандартного нормального распределения случайной величины E_S:

$$f_S(\varepsilon_S) = \frac{1}{\sqrt{2\pi}} e^{-\varepsilon_S^2}. \qquad (10)$$

Раскрывая в (9) с учетом соотношений (6), (10) выражения для плотностей вероятностей, получаем:

$$f(\varepsilon) = \frac{1}{\Delta^2 \sqrt{2\pi}} \int_0^{\infty} e^{-\frac{\sigma_\varepsilon^2}{2\Delta^2}} e^{-\frac{\varepsilon^2}{2\sigma^2}} d\sigma_\varepsilon = \frac{1}{\Delta^2 \sqrt{2\pi}} \int_0^{\infty} e^{-\delta\sigma_\varepsilon^2 - \beta/\sigma_\varepsilon^2} d\sigma_\varepsilon, \qquad (11)$$

где $\delta = 1/2\Delta^2$, $\beta = \varepsilon^2/2$.

Интеграл (11) является частным случаем табличного интеграла [12 , формула 3.478.4], нахождение которого требует вычисления цилиндрических функций. Опуская промежуточные вычисления, приводим окончательный результат:

$$f(\varepsilon) = \frac{1}{2\Delta} e^{-\frac{|\varepsilon|}{\Delta}}, \qquad (12)$$

ISSN 1562-9945

что соответствует выражению для плотности случайной величины, распределенной по закону Лапласа. В отличие от выражения (7), формула (12) уже не содержит неопределенности, обусловленной наличием случайного параметра. Практическое подтверждение лапласового распределения погрешности результатов групповой экспертизы получено в [12, 13].

Выводы

Предложена модель распределение погрешностей экспертно-аналитических оценок, рассчитываемых по данным групповой экспертизы, представленная нормальным законом распределения со случайным параметром – СКО. При достаточно общих предположениях о факторах, влияющих на формирование погрешностей, показана возможность описания распределения этого параметра законом Рэлея, что обуславливает трансформацию исходного нормального распределения со случайным параметром в закон Лапласа с полностью детерминированным описанием.

ЛИТЕРАТУРА

1. Сетлак Г. Интеллектуальные системы поддержки принятия решений. – К.: Логос, 2004. – 251 с.
2. Гнатієнко Г.М. Експертні технології прийняття рішень / Г.М. Гнатієнко, В.Є.Снитюк: Монографія. – К.: ТОВ «Маклаут», – 2008. – 444 с.
3. Информационные технологии организации бизнеса / [Карпенко С.В., Иванченко Е.В., Корченко А.А., Казмирчук С.В.]. – К.: Изд-во Национального авиационного ун-та, 2012. – 306 с.
4. Петренко С.А. Управление информационными рисками. Экономически оправданная безопасность. / С.А. Петренко, С.В. Симонов – М.: Компания АйТи; ДМК Пресс, 2004. – 384 с Павлов А.А. Модифицированный метод анализа иерархий (версии 2,3) / А.А.Павлов, А.А.Иванова, А.С.Штанькевич, А.П.Федотов // Вісник НТУУ «КПІ» – К.: «ВЕК+», 2010 – № 50. – С.3-13. – (Серія «Інформатика, управління та обчислювальна техніка»)
5. Гаврилова Т.А. Извлечение и структурирование знаний для экспертных систем /Т.А. Гаврилова, К.Р. Червинская – М.: Радио и связь, 1992.–200 с.
6. Поспелов Г.С. Искусственный интеллект – основа новой информационной технологии / Г.С.Поспелов – М.: Наука, 1988. – 280 с.
7. Ивашко В.Г. Экспертные системы и некоторые проблемы их интеллектуализации / В.Г.Ивашко, В.К.Финн // Семиотика и информатика, М.: ВИНИТИ, 1986. - №27. – С.25-61.
8. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и её инженерные приложения / Е.С. Вентцель, Л.А. Овчаров – М.: Наука, 1988. – 480 с.

9. Пугачев В.С. Теория вероятностей и математическая статистика / В.С. Пугачев – М.: Наука, 1979. – 496 с.

10.Архипов А.Е. О моделировании некоторых типов случайных последовательностей / А.Е. Архипов // Вестник Киев. политехн. ин-та – Вып. 12. – К.:1988 – С. 39-44.

11.Градштейн И.С. Таблицы интегралов, сумм, рядов и произведений / И.С. Градштейн, И.М. Рыжик – М.: ГИФМЛ, 1971.

12.Носок С.О. Методи обробки експертних даних в задачі автоматизації профвідбору : дис. ... канд. техн. наук за спец. 05.13.07 Автоматизація технологічних процесів: захищена 22.10.2007 / Носок Світлана Олександрівна.– К., 2007. – 160 с.

13.Архипов А.Е. Анализ и обработка данных артикуляционных испытаний / А.Е.Архипов, Е.А.Архипова // Захист інформації.– 2012.– №4 (57), – С.34 – 42.

 ISSN 1562-9945