

Н.Г. Топольский, И.М. Тетерин, Д.С. Шапошник, Н.Ю. Рыженко, А.А. Рыженко
(Академия ГПС МЧС России; e-mail: litloc@rambler.ru)

ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ ПРИ ДИАГНОСТИРОВАНИИ УРОВНЯ ПОДГОТОВЛЕННОСТИ ПЕРСОНАЛА ОРГАНИЗАЦИЙ К ДЕЙСТВИЯМ ПРИ ЧС

Приводятся результаты анализа методов обработки статистических данных при диагностировании уровня подготовленности персонала организаций к действиям в ЧС. Учтён дополнительный оценочный критерий, позволяющий обосновать степень адекватности заданий. Материал может быть полезен при решении проблем техносферной безопасности.

Ключевые слова: статистика, массивы данных, оценка, диагностика.

N.G. Topolsky, I.M. Teterin, D.S. Shaposhnik, N.Yu. Ryzhenko, A.A. Ryzhenko

PROCESSING OF STATISTICAL DATA WHEN DIAGNOSING LEVEL OF READINESS PERSONNEL ORGANIZATIONS TO ACTIONS IN CASE EMERGENCY

The results of the analysis methods of processing statistical data when diagnosing level of readiness personnel organizations to actions in case emergency. The additional estimated criterion allowing proving degree of adequacy tasks is considered. The material can be helpful in solving the problems of technosphere safety.

Key words: statistics, data files, assessment, diagnostics.

Статья поступила в редакцию Интернет-журнала 13 февраля 2015 г.

Введение

Аппаратные, программные, социальные, экономические и прочие сложные системы контроля и диагностики состояния, как правило, поддерживают двухкритериальную систему диагностики и оценки. В качестве корректного состояния выступает балл ИСТИНА или "1", противоположной оценкой является балл ЛОЖЬ или "0". Данная система вполне оправдана для технических комплексов или программной среды в связи с введенной стандартизацией при кодировании в бинарной системе счисления [1-3, 5]. Для других сфер диагностической деятельности данный метод может комбинироваться в последовательности.

Например, при оценке действий персонала в случае ЧС возможно множество вариантов, приводящих к положительному решению. При эвакуации из здания можно использовать несколько путей, причем на каждом разветвлении появляются дополнительные решения. В результате полученное дерево решений имеет множество вероятных состояний с положительным или отрицательным результатом. Аналогично можно рассматривать последовательность действий при тушении очагов возгорания подручными средствами и пр.

В статье представлена выборка статистических методов обработки данных двухкритериальной оценочной системы, позволяющая проводить оценку с учётом специфики диагностирования подготовленности к действиям в ЧС на основе временной шкалы и бинарного дерева решений.

Экспресс метод

Этапы экспресс метода [6]:

- построение матрицы **двухкритериальных заданий (ДЗ)** на основе сформированного дерева решений в определенный момент времени;

- определение общей меры трудности ДЗ и уровня трудности каждого ДЗ, расчёт коэффициента корреляции. Представляет собой обоснованные весовые коэффициенты. Например, при выборе направления эвакуации есть более прерогативные, менее загруженные и т.д.;

- нахождение вариации, дисперсии баллов ДЗ и дифференцирующей способности, определение коэффициентов надежности. Например, проводится оценка и вероятность применения подручных средств для тушения очага воспламенения в определенный момент времени;

- проведение рейтинга испытуемых (объект испытания определенной среды за отведенный интервал времени). Например, последовательность действий определенного сотрудника при задымлении закрытого помещения.

Исходные для дальнейшего анализа результаты прохождения ДЗ хранятся в форме прямоугольных матриц. В каждой ячейке содержится результат оценки действий персонала в определенный момент времени на основе дерева решений. При этом учтено условие: для обоснования критерия достоверности, число (N) испытуемых должно быть больше числа (k) ДЗ. Требуемое для оценки соотношение числа испытуемых к числу заданий – не менее пяти. В получаемой матрице результатов подсчитывается количество баллов ИСТИНА (R_j) и ЛОЖЬ (W_j), затем определяется доля баллов ИСТИНА каждого испытуемого (p_j):

$$p_j = \frac{R_j}{N},$$

где N – число испытуемых.

Параллельно с показателем трудности ДЗ задается противоположная статистика – доля результатов баллов ЛОЖЬ (q_j), вычисляемая из отношения общего числа баллов ЛОЖЬ к числу испытуемых (N):

$$q_j = \frac{W_j}{N}.$$

Для приведенных показателей необходимым условием является показатель целостности приведенных отношений:

$$p_j + q_j = 1.$$

На основании полученных данных определяются потенциал трудности заданий $\left(\frac{q_j}{p_j}\right)$ и показатель (логит) трудности задания $\beta_j = \ln\left(\frac{q_j}{p_j}\right)$.

Дополнительно вводится логарифмическая оценка (логит) уровня подготовленности испытуемых

$$\theta_i = \ln \frac{p_i}{q_i},$$

где p_i – доля показателя ЛОЖЬ испытуемого: $p_i = \frac{Y_i}{k}$;

Y_i – число баллов ИСТИНА испытуемого i ;

k – общее число заданий.

Полученный средний арифметический балл испытуемых определяет исходные данные для дальнейшего анализа:

$$M = \frac{\sum Y_i}{N}.$$

В итоговой матрице результатов проводится два упорядочения:

- по испытуемым, от самого успешного по нисходящей сумме баллов;
- для ДЗ, от самого легкого (имеется наибольшее число баллов ИСТИНА).

На основе полученных данных, результаты ДЗ разделяются по методу *F. Baker* (табл. 1).

Таблица 1

Уровни трудности заданий по методу *F. Baker*

Градации трудности	Мера
Очень трудные задания	$\beta_j > 2,6$
Трудные задания	$1,5 < \beta_j < 2,59$
Задания среднего уровня трудности	$-1,49 < \beta_j < 1,49$
Легкие задания	$-2,59 < \beta_j < -1,5$
Очень легкие задания	$\beta_j < -2,6$

Коэффициент корреляции

Коррелируемость ДЗ проверяется коэффициентом корреляции r_{xy} (r – коэффициент корреляции Пирсона). Формируется два вектор-столбца, один из которых – задание (X_j), другой – критерий (Y_i). Между значениями устанавливается мера связи результатов испытуемых по j заданию с суммой баллов испытуемых по итоговому ДЗ, строится вспомогательная таблица:

- значения баллов, полученных испытуемыми в j задании и сумма этих баллов $\sum X_j$;
- аналогично баллы (Y_i) и их сумма $\sum Y_i$;
- произведения баллов каждого испытуемого по j заданию (X_j) и по сумме баллов (Y_i): $X_j Y_i$ и их сумма $\sum X_j Y_i$;
- квадраты значений X_j и Y_i и их суммы соответственно $\sum X_j^2$ и $\sum Y_i^2$.

Производится расчёт коэффициента корреляции:

- сумма квадратов отклонений баллов испытуемых от среднего арифметического балла

$$SS_x = \sum X_j^2 - \frac{(\sum X_j)^2}{N};$$

- сумма квадратов отклонений баллов испытуемых от среднего арифметического балла по итоговому ДЗ (SS_y)

$$SS_y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N};$$

- скорректированная, на средние значения, сумма попарных произведений X и Y , по формуле

$$SP_{xy} = \sum XY - \frac{(\sum X)(\sum Y)}{N};$$

$\sum XY$ – сумма произведений баллов каждого испытуемого по j заданию Y баллу ДЗ;

- коэффициент корреляции

$$r_{xy} = \frac{SP_{xy}}{\sqrt{SS_x \cdot SS_y}}.$$

Чем выше значения показателя r_{xy} , тем больше вероятность задания на включение в итоговые ДЗ. Нижняя граница r_{xy} равна 0,3. При значениях в диапазоне $0,2 < r_{xy} < 0,5$ дополнительно используются другие критерии: мера корреляции задания, факторная чистота и т.д. Нулевая корреляция, а также задания с отрицательными значениями r_{xy} , устраняются.

При наличии больших объёмов обрабатываемых данных используется *point-biserial* коэффициент корреляции Пирсона:

$$r_{pb} = \frac{M_1 - M_2}{s_y} \sqrt{\frac{n_0 n_1}{n(n-1)}},$$

где M_1 – среднее арифметическое получивших по заданию один балл;
 M_2 – среднее арифметическое получивших по заданию ноль баллов;
 s_y – отклонение баллов от среднего арифметического:

$$s_y^2 = \frac{SS_y}{n-1}, \quad s_y = \sqrt{\frac{SS_y}{n-1}}, \quad SS_y = \sum (Y_i - M)^2,$$

n – количество испытуемых;

n_1 – число испытуемых, получивших в задании один балл;

n_0 – число испытуемых, получивших в задании ноль баллов.

На практике при сравнении r_{pb} и r_{xy} подтверждается сходство полученных значений и практическая достаточность использования одной из формул.

Надёжность результатов двухкритериальных заданий

Первый метод – метод *L.L. Guttman*: чем больше ошибок (инвертированных элементов, стоящих не на своих местах), тем хуже структура ДЗ, тем ниже надёжность результатов:

$$r_g = 1 - \frac{\sum e}{Nk},$$

где r_g – коэффициент структурированности результатов испытуемых данной группы;

$\sum e$ – количество ошибочных элементов в профилях всех испытуемых;

N – число испытуемых в матрице данных;

k – число заданий в ДЗ.

Второй метод – метод деления результатов ДЗ. По данным итоговой матрицы суммируются баллы в нечетных и в четных позициях. Полученные результаты образуют два вектора. Определяются произведение векторов и квадрат отклонения каждого из них. Рассчитывается коэффициент корреляции $r_{k_1k_2}$ (коэффициент корреляции Пирсона). Затем применяется коррекция *Spearman-Brown*, где n – дополнительный множитель, k – коэффициент корреляции:

$$r = \frac{n r_{k_1k_2}}{1 + (n-1) r_{k_1k_2}}.$$

Если полученное значение менее 0,8, то задание считается ненадёжным.

Третий метод – метод *Kuder G.F., Richardson M.W.*:

$$r_{KR-20} = \frac{k}{k-1} \left(1 - \frac{\sum p_j q_j}{s_x^2} \right),$$

где r_{KR-20} – коэффициент надёжности;

k – число заданий;

$\sum p_j q_j$ – сумма дисперсий ДЗ;

s_x^2 – дисперсия полученных баллов.

В существующих разработках компьютерных программ также используется дополнительный коэффициент альфа (α), на тот случай, когда оценки не только 1 и 0, но и другие, например, 2, 3 и так далее [4, 7]:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_j^2}{s_z^2} \right),$$

где k – количество ДЗ;

S_j^2 и S_z^2 – дисперсии баллов $\left(S_z^2 = \frac{SS_x}{N} \right)$.

Аналогично предыдущим методам, в качестве нижней границы допустимой надёжности измерения используется значение 0,8.

Подводя итог проделанной работы, предлагается структура анализа многовариантных статистических данных для представления способов анализа результатов (рис. 1) и обобщенный алгоритм реализации анализа (рис. 2).



Рис. 1. Структура анализа многовариантных статистических данных

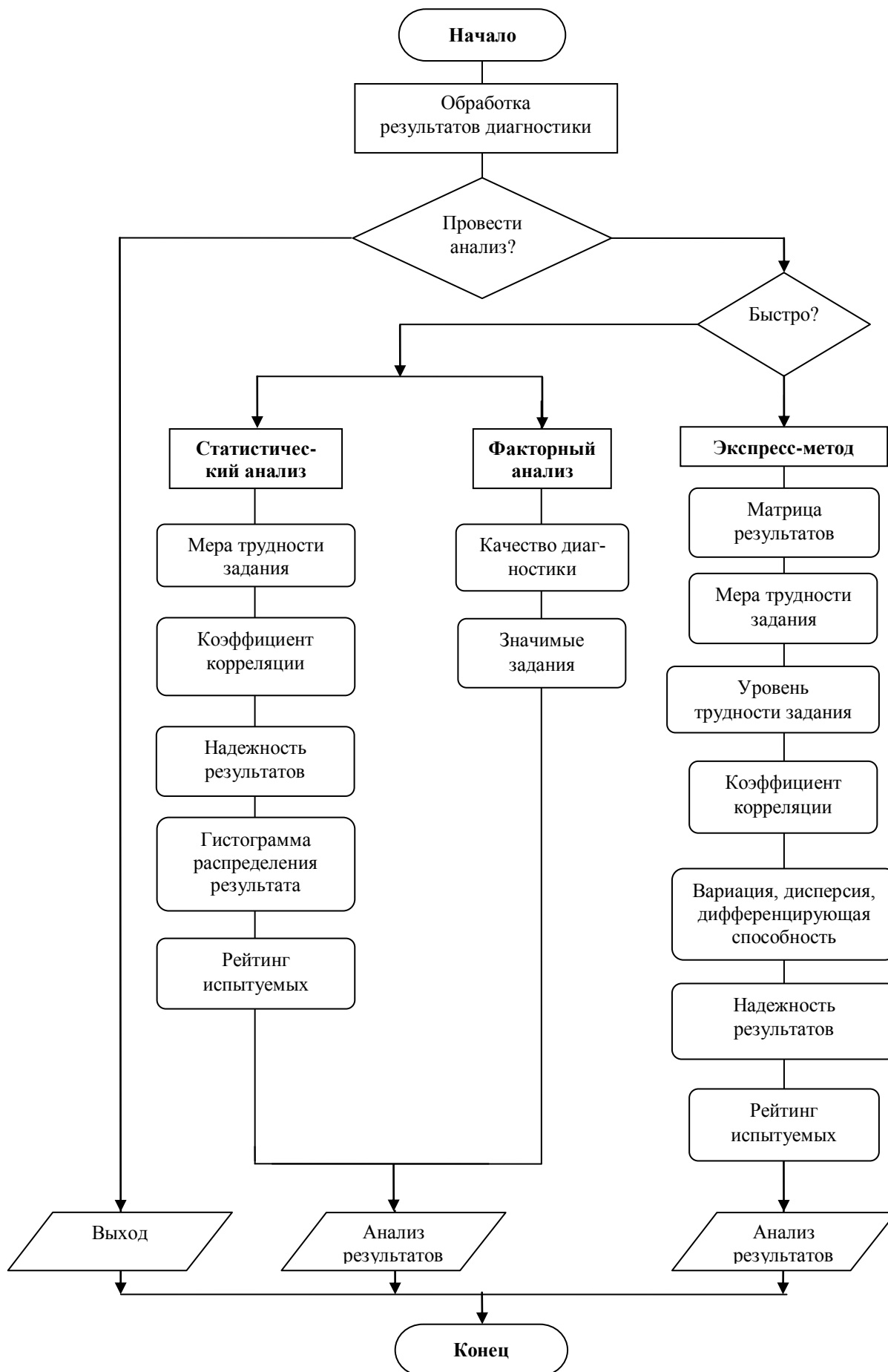


Рис. 2. Блок-схема алгоритма анализа результатов диагностики

Вывод

В основе принятого метода оценки качества результатов статистической обработки данных лежит идея рейтинга. Как правило, критерий имеет либо бинарное отношение, либо разбивается по весовым критериям на группы по балльно-лингвистической оценке. Для дальнейших исследований авторами принят рейтинг по принципу успеваемости, где испытуемых делят на три группы: успешные, подготовленные и неуспешные. Данная форма предоставляет возможность гибкой системы диагностирования в условиях постоянно меняющихся внешних условий, что является необходимым требованием развития систем поддержки управления при подготовке персонала организаций к действиям при ЧС и пожарах.

Литература

1. **Рыженко А.А., Рыженко Н.Ю.** Модель трехкритериального анализа многовариантных статистических данных // Информационные технологии в региональном развитии. Вып. IX. Апатиты, 2009. С. 62-66.
2. **Рыженко А.А.** Фасетный метод анализа статистических данных по пожарам на территории РФ // Матер. 3-й междунар. науч.-практ. конф. молодых учёных и специалистов "Проблемы техносферной безопасности – 2014". М.: Академия ГПС МЧС России, 2014. С. 297-299.
3. **Рыженко А.А.** Инструментальная система поддержки контроля уровня знаний персонала промышленного предприятия // Прикладные проблемы управления макросистемами. Т. 39. М.: Книжный дом "ЛИБРОКОМ", 2009. С. 368-375.
4. **Рыженко А.А., Яковлев С.Ю., Шемякин А.С.** Информационные системы поддержки контроля управления персоналом промышленных предприятий в тестовой форме // Св-во о регистрации электронного ресурса в ОФЭРНиО ИНИМ РАО № 18068 от 28.03.12.
5. **Рыженко А.А., Топольский Н.Г., Хабибулин Р.Ш.** Алгоритм обработки показателей системы взаимодействия комплексов безопасности и жизнеобеспечения в пределах одного объекта // XII Всероссийское совещание по проблемам управления ВСПУ-2014. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2014. С. 8230-8241.
6. **Неровных А.Н., Калайдов А.Н., Миколайчук М.М.** Математическая модель оценки качества подготовки специалистов // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. 2010. № 1. С. 115-125.
7. **Молого О.О.** Анализ способов управления кадровыми системами с использованием математических моделей // Пожары и чрезвычайные ситуации: предотвращение, ликвидация. 2011. № 4. С. 47-52.