

МЕТОД АВТОМАТИЗАЦИИ ОЦЕНКИ КАЧЕСТВА РАБОТЫ ЭКСПЕРТОВ РЕЗУЛЬТАТОВ ИТОГОВЫХ ИСПЫТАНИЙ В ФОРМЕ ЕГЭ

Щербинин Т. А., аспирант; Гранков М. В., доц., к.т.н.

(Донской государственной технической университет, г. Ростов-на-Дону, Россия)

Введение. С 2009-го года единственной формой выпускных экзаменов в школе и одновременно вступительных экзаменов в ВУЗы стал экзамен в форме ЕГЭ (Единый Государственный Экзамен).

На данном этапе в обществе сложилось неоднозначное отношение к данной форме экзамена. В связи с этим задания для экзаменов по предметам постоянно корректируются и претерпевают изменения.

По большинству предметов задания делятся на три группы, в зависимости от их сложности. При этом ответы на задания группы А (с выбором варианта ответа) и группы В (краткий ответ) заполняются на специальном бланке в строгом соответствии с представленным шрифтом для заполнения, так как проверка осуществляется на компьютере. Задания группы С, требующие развернутый ответ, записываются на специальный бланк и проверяются экспертами региональной аттестационной комиссии. Стоит отметить, что нередко случаи некорректного выставления баллов экспертами при проверке заданий части С. В связи с этим конечная оценка, полученная экзаменуемым, зависит не только от правильности выполненных на экзамене заданий, но и от компетенции проверяющего эксперта, который может допустить ряд ошибок при проверке работы. Выпускник также может обжаловать результаты оценивания в специальной комиссии.

Таким образом, важной задачей становится проблема формирования экспертной комиссии, проверяющей задания части С.

Ошибки при проверке экзаменационных работ. Процедура проверки экспертом нормируется специальными инструкциями «Федеральный институт педагогических измерений» (ФИПИ), которые содержат подробные схемы (алгоритмы) оценки каждой задачи из каждого варианта задания. Для реализации оценивания каждому эксперту предоставляется сканкопии реальных работ выпускников.

Для сокращения влияния ошибок эксперта на результат оценивания для каждого конкретного задания области «С» предусмотрена процедура проверки двумя экспертами этой или другой региональной предметной комиссии. Если оценки этих экспертов имеют несущественные различия, то учитывается большая из них. Для каждого предмета определен свой критерий несущественности различия оценок. Если же оценки экспертов недостаточно близки согласно определенному для этого предмета критерию, то данное задание автоматически отправляется на проверку третьему эксперту, решение которого считается окончательным.

Методика оценки. Разработанная методика оценивает качество работы экспертов по количеству проверенных задач. Производится анализ числовых показателей по результатам проверки экспертами этих задач.

Были использованы три группы для оценивания качества экспертов: «Низкая (низкое качество проверки)», «Средняя (среднее качество проверки)», «Высокая (высокое качество проверки)». В результате множественных экспериментов на различных экзаменационных комиссиях пришли к выводу, что основными показателями оценки качества работы являются частота корректировок при 3-х проверках и удельный балл оценивания, который играет важную роль если показатели частоты корректировок экспертов близки между собой.

Частоту корректировок баллов при 3-х проверках d_i^{e3} для i -го эксперта, будем рассчитывать по формуле (1):

$$d_i^{e3} = \frac{t_3}{m_i}, \quad (1)$$

где t_3 – количество ошибок, допущенных при 3-х проверках;

m_i – количество заданий, проверенных i -м экспертом [1, 2].

Альтернативный балл ($xalt3_i^k$) для i -го эксперта для задания k , в котором была 3-я проверка, равняется баллу, который был выставлен 3-м экспертом:

$$xalt3_i^k = x_3^k, \quad (2)$$

где x_3^k – оценка, выставленная 3-м экспертом за k -е задание (может отсутствовать).

Расчет абсолютного значения отклонения оценивания ($xabs3_i^k$) для i -го эксперта в k -ом задании с 3-й проверкой, найдем по формуле (3):

$$xabs3_i^k = |xalt3_i^k - x_i^k|, \quad (3)$$

где x_i^k – оценка, выставленная i -м экспертом за k -е задание;

$xalt3_i^k$ – альтернативный балл для k -го задания по формуле (2).

Удельное абсолютное значение отклонения в 3-х проверках ($xabst3_i^k$) найдем по формуле (4):

$$xabst3_i^k = \frac{\sum_{k=1}^{m_i^3} xabs3_i^k}{m_i}, \quad (4)$$

где $xabs3_i^k$ – отклонение значение отклонения оценивания i -м экспертом задания с номеров k по формуле (4);

m_i^3 – количество заданий с 3-ми проверками у i -ого эксперта;

m_i – количество заданий, проверенных i -м экспертом.

Метод кластеризации. В качестве метода кластеризации оцененных экспертов, то есть объединения их в группы по полученным результатам, был использован алгоритм Fuzzy C-Means (FCM) [2]. Этот алгоритм чаще всего применяется для нечеткого группирования данных. Он формирует группы, прототипы которых представляются точками в пространстве данных. Для выполнения алгоритма на некотором множестве данных X необходимо выбрать количество групп s , степень нечеткости n , параметр ε в критерии останова алгоритма, а также случайным образом инициировать матрицу принадлежности объектов к той или иной группе

$$U^{(0)} \in Z_i,$$

где i – количество групп, и вектор прототипов $V^{(0)}$.

В качестве условия завершения алгоритма (критерия останова) рассматривается достаточно малое изменение значений элементов матрицы U , то есть $\|xalt3_i^k - x_i^k\| < \varepsilon$, где ε – заданная константа [3].

Начальные центры кластеров распределяются равномерно по центрам отрезков от минимального значения нормированного параметра до максимального. Центры кластеров выбираются таким образом, чтобы минимизировать расстояние от центра до группы

объектов и максимизировать расстояние между центрами различных кластеров. Всего количество кластеров принимается равным трем и равняется количеству групп оценивания экспертов (низкое качество проверки, среднее качество проверки и высокое качество проверки).

Начальное значение центра i -го кластера ($centr_i$) находится по формуле:

$$centr_i = \min(par) + i * \frac{\max(par) - \min(par)}{2 * count}$$

где $\min(par)$ — минимальное значение нормированного параметра из выборки;
 $\max(par)$ — максимальное значение нормированного параметра из выборки;
 $count$ — общее количество кластеров.

В качестве меры удаленности была использована взвешенная метрика Эвклида, согласно которой расстояние ($dist$) от точки до кластера находится по формуле:

$$dist = \sqrt{k_1(x_c - x)^2 + k_2(y_c - y)^2}$$

где x_c, y_c — значение координат центра кластера;

x, y — значение координат текущего эксперта;

k_1, k_2 — значения весовых коэффициентов оцениваемых параметров [2].

Анализ полученных результатов. Рассмотрим анализ работы экспертов на примере комиссии по химии.

Очевидно, что во всякой работе с выполнением однотипных операций вырастает усталость и вероятность ошибок, поэтому все параметры оценивания будем соотносить к количеству проверенных задач.

Так как корректировки считаются более важными, чем удельный балл отклонения, было принято, что их оптимальное отношение 1 к 3.

Рассмотрим пример распределения экспертов на группы по качеству проведенных проверок. Используемые в таблицах сокращения описаны в таблице 1.

Таблица 1 Используемые сокращения

N	Порядковый номер эксперта
PAR	Значение частоты корректировок в 3-х проверках (к общему числу заданий)
PAR2	Значение удельной величины абсолютного балла отклонения
PAR_N	Нормированное значение PAR
PAR2_N	Нормированное значение PAR2
G (%)	Принадлежность к кластеру хорошего кластера проверки
M (%)	Принадлежность к кластеру среднего качества проверки
B (%)	Принадлежность к кластеру плохого качества проверки
REIT	Рейтинг эксперта
BEL	Результатирующая принадлежность (Рассчитывается в зависимости от большей принадлежности): G (Good) хорошее качество проверки, M(Medium) среднее качество проверки, B(Bad) плохое качество проверки.

Таблица 2 Оценка работы экспертов по взвешенной метрике Эвклида

N	PAR	PAR2	PAR_N	PAR2_N	G	M	B	DIST_0	BEL
1	0,84%	0,01145	0	0	93,71%	4,63%	1,66%	1	G
2	0,87%	0,013044	0,015	0,038	95,02%	3,69%	1,29%	0,983	G
3	1,01%	0,017998	0,084	0,155	99,28%	0,55%	0,17%	0,908	G
4	1,05%	0,016554	0,104	0,12	99,77%	0,18%	0,05%	0,894	G
5	1,16%	0,018813	0,155	0,174	99,21%	0,62%	0,17%	0,843	G
6	1,20%	0,022202	0,177	0,254	96,70%	2,64%	0,66%	0,814	G
7	1,31%	0,018868	0,228	0,175	90,63%	7,63%	1,73%	0,779	G
8	1,36%	0,025989	0,252	0,343	79,99%	16,72%	3,30%	0,738	G
9	1,81%	0,026499	0,475	0,355	9,22%	86,22%	4,56%	0,54	M

10	2,01%	0,040189	0,571	0,679	0,73%	98,13%	1,14%	0,418	М
11	2,06%	0,034921	0,597	0,554	0,42%	98,85%	0,74%	0,408	М
12	2,09%	0,038239	0,608	0,632	0,90%	97,22%	1,88%	0,389	М
13	2,62%	0,043071	0,87	0,747	1,31%	8,17%	90,52%	0,15	В
14	2,79%	0,041783	0,95	0,716	0,28%	1,34%	98,38%	0,108	В
15	2,89%	0,048889	1	0,884	0,25%	1,03%	98,72%	0,04	В
16	2,89%	0,053806	0,999	1	0,59%	2,38%	97,03%	0,001	В

На рисунке 1 представлены графики распределения экспертов по группам. Эксперты с плохим качеством проверки обозначены квадратами, с средним – треугольниками, с высоким – ромбами. Черными точками на графиках обозначены центры кластеров. Ранжирование экспертов происходит по параметру REIT.

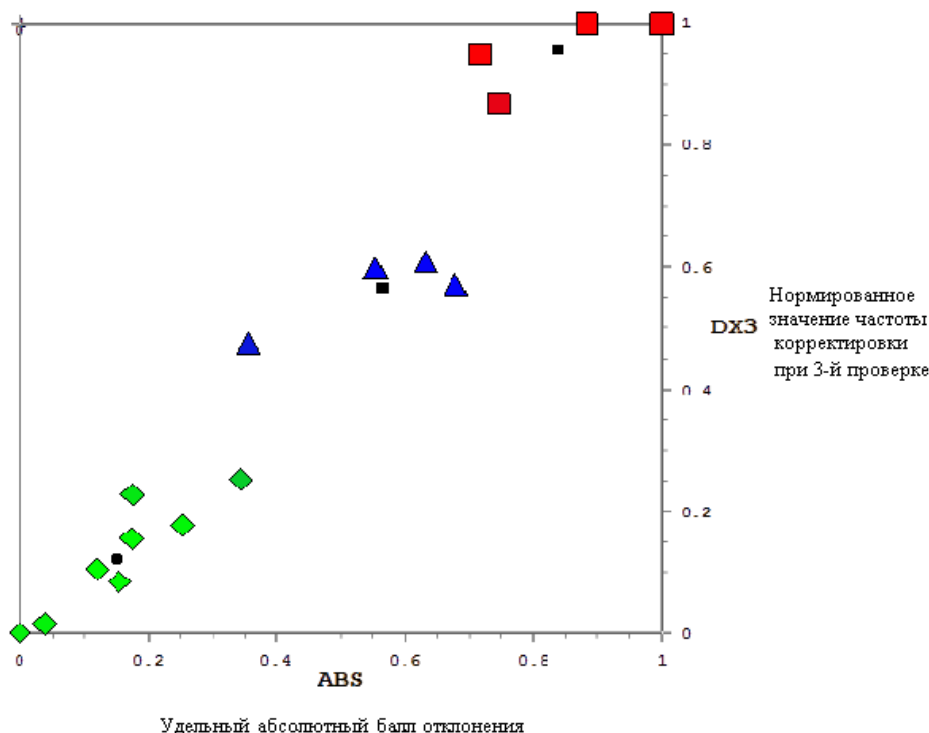


Рисунок 1 - График распределения экспертов на группы

Выводы. Рассмотренная методика позволяет оценить качество работы экспертов предметных комиссий. Опираясь на полученное распределение экспертов по группам, можно улучшить работу региональной экзаменационной комиссии. Предлагается по возможности формировать комиссию на следующий год в соответствии с полученным распределением, отказываясь от экспертов, которые попали в группу низкого качества проверки. Экспертам со средним качеством проверки следует дать рекомендации по поводу их работы, позволяющие улучшить их показатели проверки на следующий год. Качественную проверку заданий осуществляют эксперты из группы с высоким качеством, их рекомендуется приглашать в комиссию проверки. Эксперты из группы с хорошим качеством проверки оценивают задания наилучшим образом и их рекомендуется приглашать в комиссию на следующий год.

Перечень ссылок

1. Гранков М.В., Щербинин Т.А.. Оценка параметров работы экспертов предметной комиссии по проверке экзамена в форме ЕГЭ. Системный анализ, управление и обработка информации: Труды 5-го Международного семинара (п. Дивноморское, 2 – 6 октября 2014 г.) / Под общ. ред. Р.А. Нейдорфа. – Ростов-н/Д: ДГТУ, 2014. – С. 282-285.

2. Рутковский Л. Методы и технологии искусственного интеллекта – Горячая Линия – Телеком, 2010. – 520 с.