УДК 37.016:51

МЕТОД СТАТИСТИЧЕСКОЙ ОЦЕНКИ ОТНОСИТЕЛЬНОЙ СЛОЖНОСТИ ОЛИМПИАДНЫХ И ТЕСТОВЫХ ЗАДАЧ

METHOD OF STATISTICAL ESTIMATION OF ACADEMIC COMPETITION AND TESTING TASKS RELATIVE COMPLEXITY

Лазарев В.А., Хайбуллин Р.Я.
ФГБОУ ВПО «Уфимский государственный нефтяной технический университет», г. Уфа, Российская Федерация

V.A. Lazarev, R.Y. Khaybullin
FSBEI HPE "Ufa State Petroleum Technological University", Ufa, the Russian Federation
e-mail: lazva@mail.ru, khayrya@mail.ru

Аннотация. Данная статья посвящена вопросу оценки относительной трудности задач при проведении предметных олимпиад среди школьников и студентов. Как показывает опыт, выставляемые организаторами оценки (в баллах) за правильные решения задач зачастую не соответствуют наблюдаемой после проверки статистике. Основные трудности здесь связаны с практической невозможностью подвести какую-либо формальную базу, сформулировать универсальный критерий для такой оценки. Даже для учебных школьных задач коэффициент корреляции между экспертными и статистическими оценками может оказаться меньше 0,5. Тем не менее, адекватное соотношение баллов и трудности решения задач способствует более справедливому ранжированию участников по уровню их знаний и умений, что и является основной целью любого теста, олимпиады, конкурсного испытания. В данной работе предлагается метод вычисления наиболее адекватных оценок. Он основывается на результатах проверки работ уже проведенной олимпиады и поэтому результаты носят апостериорный характер. Сформулированы два критерия адекватности оценки сложности и для каждого из них приведены соответствующие математические задачи. Проведенный анализ конкретных олимпиад позволяет сделать выводы о правильности относительной оценки задач, наличия незначимых задач среди предложенных, роли той или иной задачи для достижения цели мероприятия и т.д. В частности, часто на олимпиадах предлагается избыточное количество задач, причем сложные задачи, как правило, недооцениваются, а легкие наоборот переоцениваются организаторами. Предложенные методы применимы для любых конкурсных испытаний и тестов по любым предметам, как для школьников, так и студентов.

Abstract. This article deals with the problem of relative complexity problems estimation at the thematic competitions among pupils and students. The experience shows that the grades (in points) given by organizers for correct solving of problems do not often correspond the statistics given after inspection. The main difficulties here are connected with the problem that it's impossible to give any formal ground, formulate a multipurpose criterion for such estimation. Even for academic school problems the correlation ratio between expert and statistical estimation may be less than 0.5. Nevertheless, adequate interrelationship of points and difficulties of solving mathematical problems facilitates more fair arranging of participants taking into account their knowledge and skills level, which is the main purpose of any test, academic competition, contest. The present work suggests the calculation method of the most adequate estimation. It is based on the results of the works of the conducted academic competition and therefore the results are considered to be of posteriori character. Two criteria of adequate estimation of complexity are formulated and relevant mathematic problems for each of them are given. The conducted analysis for specific academic competitions let us make conclusions concerning the correctness of problems relative estimation, the occurrence of insignificant problems among given etc. In particular, abundant number of problems are often offered at the academic competitions, but as a rule, complex problems are underestimated and easy ones are overestimated by organizers. Suggested methods are applied for any contests and tests on any subjects, both for pupils and students.

Ключевые слова: оценка сложности, предметные олимпиады, апостериорный анализ, функция Лагранжа.

Key words: complexity estimation, thematic competitions, posteriori analysis, Lagrange function.

Введение

При подготовке любого конкурсного задания, состоящего более чем из одной задачи, возникает проблема оценки трудности каждой отдельной задачи [1]. Проще говоря, речь идет о присваивании каждой задаче некого номинального балла за полное безукоризненное решение. Этот балл в данной статье мы будем называть априорной оценкой трудности задачи (в литературе также употребляется термин «директивная» трудность). Адекватное соотношение баллов и трудности решения задач способствует справедливому ранжированию участников по уровню их знаний и умений.

Оценить относительную сложность олимпиадной задачи весьма затруднительно, ввиду практической невозможности подвести какую-либо формальную базу, сформулировать универсальный критерий. Конечно, можно как-то учитывать количество этапов решения, их сложность, уровень используемых в решении теорем и так далее. Однако, проблем здесь гораздо больше, чем решений, даже для учебных задач [2]. Чаще эти вопросы решаются экспертным методом – члены оргкомитета путем обсуждения, сравнения мнений принимают решение о номинале каждой задачи. И, несмотря на высокую квалификацию организаторов, очень часто случаются неадекватные оценки. То есть при проверке работ оказывается, что более легкую задачу решает меньшее, чем можно было бы ожидать, число участников, более трудную – наоборот, большее. Даже для учебных школьных задач коэффициент корреляции между экспертными и статистическими оценками может оказаться меньше 0,5 [3]. Этому можно привести какие-то объяснения, но это не входит в задачи данной статьи.

Данная статья посвящена апостериорному анализу сложности задач. Под этим понимается выяснение вопроса, насколько удачно были выбраны априорные баллы для оценки задач и как они должны были выглядеть в более удачном варианте.

Постановка задачи и обозначения

Определимся более четко с постановкой задачи и обозначениями. Будем считать, что на олимпиаде предложено N задач, каждая из которых оценивается в некоторый балл $B_j^{(0)}, j = 1..N$. Вектор значений $B^{(0)} = \{B_j^{(0)}\}_{j=1}^N$ будем называть вектором априорных оценок сложности задач, или просто априорной оценкой. Рассматривается задача об определении такого вектора оценок, который был бы в некотором смысле более адекватным сложности предложенных задач. Критерии адекватности будут сформулированы ниже.

Основным материалом для анализа служит сводная таблица результатов участников олимпиады, полученная после проверки работ (таблица 1).

Таблица 1. Представление результатов проверки работ

	Задание 1 ($B_1^{(0)}$)	Задание 2 ($B_2^{(0)}$)		Задание N ($B_N^{(0)}$)	Сумма
Участник 1	α_{11}	α_{12}	…	α_{1N}	W_1
Участник 2	α_{21}	α_{22}		α_{2N}	W_2
⋮					
Участник K	α_{K1}		…	α_{KN}	W_K
	V_1	V_2	…	V_N	W_0

Предполагается, что сумма баллов $\sum_{j=1}^{N} B_j^{(0)} = B_\Sigma$ в дальнейшем есть величина неизменная. Некоторый вектор оценок $B = (B_1, B_2, \dots, B_N)$ будем называть допустимым, если все компоненты B_i $(i = 1, \dots, N)$ являются неотрицательными числами и удовлетворяют условию $\sum_{j=1}^{N} B_j = B_\Sigma$. Из соображений удобства решения получающихся задач мы отказываемся от непринципиального условия целочисленности.

Далее, величины α_{ij}, $i = 1..N, j = 1..K$ равны доле от максимального балла, полученной i-м участником за j-ю задачу, где $K -$ количество участников. $W_i -$ суммарный балл i-го участника, $W_i = \sum_{j=1}^{N} \alpha_{ij} B_j$. Сумму всех набранных участниками баллов обозначим через $W_0 = \sum_{i=1}^{K} W_i$. Сумму баллов, набранных всеми участниками за j-ю задачу, обозначим через:

$$V_j = \sum_{i=1}^{K} \alpha_{ij} B_j = a_j B_j, \text{где } a_j = \sum_{i=1}^{K} \alpha_{ij}.$$

Очевидно, что каждому допустимому вектору оценок B соответствуют некоторые векторы $W = \{W_i\}_{i=1}^{K}$ и $V = \{V_j\}_{j=1}^{N}$. Рассмотрим два критерия адекватности оценок B. Первый критерий основан на том принципе, что суммы баллов $V_1, V_2, \dots, V_N$, набранных участниками за отдельные задачи, должны мало отличаться друг от друга и, соответственно, от величины W_0/N. Таким образом, мы получаем:

Критерий 1. Допустимый вектор оценок $B^{(1)}$ будем называть оптимальным, если он доставляет минимум сумме квадратов отклонений величин V_j от W_0/N:

$$R\left(B^{(1)}\right) = \min_B R(B), \tag{1}$$

где $R(B) = \sum_{j=1}^{N} (V_j - W_0/N)^2$.

Второй критерий для выбора B представляет собой характерное для факторного анализа условие максимизации дисперсии полученных участниками оценок. Обозначим через $\sigma^2(B)$ дисперсию значений вектора W. Таким образом, имеем:

Критерий 2. Допустимый вектор оценок $B^{(2)}$ будем называть оптимальным, если:

$$\sigma^2\left(B^{(2)}\right) = \max_B \sigma^2(B), \qquad (2)$$

где $\sigma^2(B) = \frac{1}{K-1}\sum_{i=1}^{K}(W_i - W_0/K)^2$.

Рассмотрим каждый из этих критериев по отдельности.

Оптимальность оценки по Критерию 1

Рассмотрим задачу о нахождении вектора $B^{(1)} = (B_1^{(1)}, \dots, B_N^{(1)})$, оптимального в смысле Критерия 1. Отметим, что зависимость от $B^{(1)}$ в формуле (1) присутствует как в V_j, так и в W_0. Таким образом, получаем задачу для $(B_1^{(1)}, \dots, B_N^{(1)})$:

$$\begin{cases} \sum\limits_{i=1}^{N}(\sum\limits_{j=1}^{K}\alpha_{ji}\, B_i^{(1)} - \frac{1}{N}\sum_{j=1}^{K}\sum\limits_{l=1}^{N}\alpha_{jl}\, B_l^{(1)})^2 = min, \\ B_1^{(1)} + \ \dots + \ B_N^{(1)} = B_\Sigma, \\ B_1^{(1)} \geq 0, \dots, B_N^{(1)} \geq 0. \end{cases}$$

Или, используя введенные выше обозначения, можно переписать задачу в более лаконичном виде:

$$\begin{cases} \sum\limits_{i=1}^{N}(a_i B_i^{(1)} - \frac{1}{N}\sum\limits_{l=1}^{N} a_l\, B_l^{(1)})^2 = min, \\ B_1^{(1)} + \ \dots + \ B_N^{(1)} = B_\Sigma, \\ B_1^{(1)} \geq 0, \dots, B_N^{(1)} \geq 0. \end{cases}$$

Эта задача на условный минимум может быть решена классическим методом множителей Лагранжа, так как условие неотрицательности $B^{(1)}$ здесь выполняется автоматически. Ответ выписывается в явном виде:

$$B_1^{(1)} = \frac{B_\Sigma}{(\frac{1}{a_1^2}+\frac{1}{a_2^2}+\cdots+\frac{1}{a_N^2})a_1^2}, \dots, B_N^{(1)} = \frac{B_\Sigma}{\left(\frac{1}{a_1^2}+\frac{1}{a_2^2}+\cdots+\frac{1}{a_N^2}\right)a_N^2}. \quad (3)$$

При этом значение функции $R\left(B^{(1)}\right)$из формулы (1) равно нулю.

Оптимальность оценки по Критерию 2

Теперь рассмотрим задачу о нахождении вектора $B^{(2)} = (B_1^{(2)}, \dots, B_N^{(2)})$, оптимального в смысле Критерия 2. Это приводит к следующей задаче на условный экстремум с линейными ограничивающими неравенствами:

$$\begin{cases} \sum_{i=1}^{K}(\sum_{j=1}^{N} \alpha_{ij}\, B_j^{(2)} - \frac{1}{K}\sum_{j=1}^{N} a_j\, B_j^{(2)})^2 = max, \\ B_1^{(1)} + \ \dots + \ B_N^{(1)} = B_\Sigma, \\ B_1^{(1)} \geq 0, \dots, B_N^{(1)} \geq 0. \end{cases}$$

Здесь условия $B_1^{(2)} \geq 0, \dots, B_N^{(2)} \geq 0$ играют существенную роль, так что для решения приходится использовать обобщенный метод множителей Лагранжа [4], либо вычислять решение приближенно тем или иным численным методом [5,6].

Результаты вычислений

Проведение данного анализа для нескольких олимпиад, в которых авторы принимали участие как организаторы, выявило некоторые устойчивые тенденции. В частности, независимо от критерия оптимальности, выяснилось, что легкие задачи организаторами, как правило, переоцениваются, а трудные – наоборот недооцениваются. Рассмотрим пример. В марте 2014 года в УГНТУ проходила олимпиада по математике среди школьников. На олимпиаде предлагалось 10 задач, которые оценивались согласно вектору $B^{(0)}$. После подведения итогов оказалось, что оптимальные по Критерию 1 и Критерию 2 значения оценок $B^{(1)}$ и $B^{(2)}$ заметно отличаются от $B^{(0)}$, (таблица 2).

Таблица 2. Сравнение априорных и апостериорных оценок

	1	2	3	4	5	6	7	8	9	10
$B^{(0)}$	3	3	5	6	6	2	6	6	6	3
$B^{(1)}$	0,2	3,2	1,4	2	1,7	0	3,8	9,5	24,8	0,3
$B^{(2)}$	0	2,9	1,5	2	0,7	0	0	4.2	35,7	0
Тема	Алгебраическая система, нелинейная замена	Тригонометрическое уравнение	Показательно-логарифмическое неравенство	Стереометрия, сечение пирамиды	Планиметрия на доказательство	Текстовая задача, приводящая к целочисленному неравенству	Параметры, геометрическое решение	Функциональное уравнение	Иррациональное уравнение, решение с помощью векторов	Теория графов.

Близкие к нулю апостериорные оценки за задачи 1, 6, 10 (все значения округлены до десятых) показывают на относительную легкость этих задач для участников олимпиады. Практически все участники решили эти задачи. Конечно, такие задачи включаются в олимпиаду с педагогическими целями, однако для ранжирования участников по силе они малоинформативны. С другой стороны, мы видим, что восьмая и девятая задачи представляли наибольшую сложность для участников, хотя и оценивались на уровне относительно легких седьмой и четвертой задач.

Необходимо отметить, что к полученным оценкам надо относиться осторожно. Так увеличение априорного балла за правильное решение повышает интерес к задаче и ее пытаются решить и решают большее число участников, что изменяет итоговую статистику, и анализ оценок приходится проводить снова. Таким образом, предложенный анализ представляет собой, по сути, только один шаг итерационного процесса.

Очевидно, что изменение вектора B влечет изменение набранных участником баллов и, следовательно, занятого места. Возникает вопрос, насколько существенна разница в занятых местах при различных векторах $B^{(0)}, B^{(1)}$ и $B^{(2)}$. Таблица 3 иллюстрирует эту разницу.

Таблица 3. Распределение первых мест при различных оценках $B^{(0)}, B^{(1)}$ и $B^{(2)}$.

$B^{(0)}$	$B^{(1)}$	$B^{(2)}$
1	1	1
2	3	2
3	6	5
4	2	3
5	8	8
6	5	12
7	4	4

В колонках таблицы 3 указаны места участников при соответствующих оценках $B^{(0)}, B^{(1)}$ и $B^{(2)}$. Например, участник, занявший третье место в олимпиаде (при векторе $B^{(0)}$), при оценках $B^{(1)}$ он занял бы шестое место, а при оценках $B^{(2)}$ был бы пятым. Как мы можем видеть, победитель один и тот же, однако в распределении остальных мест имеются существенные различия. В проигрыше оказались те участники, которые решали сложные задачи и, потратив много времени и сил, получили за них относительно мало баллов. Аналогичные выводы получены и при анализе других олимпиад различных уровней, как школьников, так и студентов. Также в качестве результата работы отметим слабую корреляцию между априорной оценкой и полученными апостериорными оценками $B^{(1)}$ и $B^{(2)}$ – 0,49 и 0,34 соответственно, что свидетельствует об актуальности исследуемой проблемы.

Выводы

На наш взгляд, полученные результаты помогают лучше понять степень сложности тех или иных тем, задач, методов для современных школьников и студентов, что важно при подготовке олимпиад и тестов в дальнейшем.

Вышеуказанные результаты мы получаем только тогда, когда олимпиада прошла и менять оценки некорректно, хотя и возможно, если это оговорено в правилах. Например, в УГНТУ имеется опыт, когда номинальный балл за решение задачи определялся только после проверки работ. Участникам выгодней всего было решить задачу, которую никто не смог решить. Даже одна такая задача могла принести победу. Это вносило в олимпиаду дополнительный соревновательный и игровой момент.

Список используемых источников

1 Аванесов В.С. Научные основы тестового контроля знаний. М.: Исследовательский центр, 1994. 135 с.

2 Бахтизин Р.Н. К вопросу об оценке освоения дисциплины по результатам тестирования // Интернет в образовании: материалы международ. науч.-практ. заоч. конф. М., 2010. С. 193-198.

3 Наймушина О.Э., Стариченко Б.Е. Многофакторная оценка сложности учебных заданий. // Образование и наука. 2010. №2(70). С. 58-70.

4 Аоки М. Введение в методы оптимизации. М.: Наука, 1977. 344 с.

5 Полак Э. Численные методы оптимизации. М.: Мир,1974. 376 с.

6 Васильев Ф.П. Численные методы решения экстремальных задач. М.: Наука, 1988. 531 с.

References

1 Avanesov V.S. Nauchnye osnovy testovogo kontrolja znanij. M.: Issledovatel'skij centr, 1994. 135 s. [in Russian].

2 Bahtizin R.N. K voprosu ob ocenke osvoenija discipliny po rezul'tatam testirovanija // Internet v obrazovanii: materialy mezhdunarod. nauch.-prakt. zaoch. konf. M., 2010.S. 193-198.

3 Najmushina O.Je., Starichenko B.E. Mnogofaktornaja ocenka slozhnosti uchebnyh zadanij. // Obrazovanie i nauka. 2010. №2(70). S. 58-70.

4 Aoki M. Vvedenie v metody optimizacii. M.: Nauka, 1977. 344 s.

5 Polak Je. Chislennye metody optimizacii. M.: Mir,1974. 376 s.

6 Vasil'ev F.P. Chislennye metody reshenija jekstremal'nyh zadach. M.: Nauka, 1988. 531 s.

Сведения об авторах

About the authors

Лазарев В.А., канд. физ.-мат. наук, доцент кафедры «Математика», ФГБОУ ВПО УГНТУ, г. Уфа, Российская Федерация.

V.A. Lazarev, Candidate of Physics and Mathematics, Associate Professor of the Chair “Mathematics”, FSBEI HPE USPTU, Ufa, the Russian Federation

Хайбуллин Р.Я., канд. техн. наук, доцент кафедры «Математика», ФГБОУ ВПО УГНТУ г. Уфа, Российская Федерация

R.Y. Khaybullin, Candidate of Engineering Sciences, Associate Professor of the Chair “Mathematics”, FSBEI HPE USPTU, Ufa, the Russian Federation

e-mail: khayrya@mail.ru