ные технологии и общество. Казанский государственный технологический университет (Казань). – 2003. – Т. 6, № 3. – С. 164–186.

3. **Тренажеры** нового поколения : особенности, возможности, перспективы / А. Г. Ройзнер // Локомотив. – 2012. – № 5. – С. 17–19.

4. **Группа** компаний Транзас [Электронный ресурс]. – URL : http://www.transas.ru/Simulation/Railway (дата обращения 10.03.2014).

5. **Тренажерный** комплекс локомотивных бригад [Электронный ресурс]. – URL : http://www.inius.ru/index.php/trenazhery-lokomotivnyx-brigad.html (дата обращения 10.03.2014).

6. **Многотиповой** тренажерный комплекс [Электронный ресурс]. – URL : http://www.rc-spectr.ru/catalog/product/category/tc/?itid=5 (дата обращения 10.03.2014).

7. **Комплексный** тренажер машинистов скоростных поездов [Электронный ресурс]. – URL : http://www.kmweg.com (дата обращения 10.03.2014).

8. **Кузница** кадров для «Сапсанов» / В. А. Владимиров // Локомотив. – 2011. – № 6. – С. 5.

9. **Тренажеры** для машинистов [Электронный ресурс]. – URL : http://zdsimulator.com.ua/trenazhery (дата обращения 10.03.2014).

10. **Многофункциональные** тренажеры [Электронный ресурс]. – URL : http://www.sydac.com.au/en/products/mobile_simulators/mobile_simulators.jsp (дата обращения 10.03.2014).

УДК 519.2

М. М. Луценко, Н. В. Шадрицева
Петербургский государственный университет путей сообщения
Императора Александра I

КООПЕРАТИВНЫЙ И ЧАСТОТНЫЙ ПОДХОДЫ К НАЗНАЧЕНИЮ ВЕСА ЗАДАНИЙ ТЕСТА

В работе сравниваются экспертный и частотный методы назначения веса заданий теста с кооперативным, рассмотренным авторами ранее [1, 2]. Определен вес Шепли заданий, найдены их значения для теста по теме «Интегральное исчисление». Сравнение проводилось по результатам тестирования студентов в 2010–2011 гг. Показано, что кооперативный вес является хорошим ориентиром при назначении веса заданий при известной структуре курса и времени на освоении его отдельных частей.

тестирование, вес заданий, кооперативные игры, вектор Шепли.

Введение

Тестирование как инструмент измерения уровня знаний давно и успешно применяется в обучении. За последние годы интерес к нему значительно возрос, что вызвано формализацией учебного процесса и развитием компьютерных средств контроля знаний. Появление математических моделей Раша, Бирнбаума и др. существенно расширили теорию и дали возможность использовать современные математические модели: теорию параметризации педагогических тестов (Item Response Theory) [3], статистические игры [4].

Одной из проблем теории тестирования является задача определения веса задания теста [3]. В большинстве тестов вес заданий считается равным, но по мере усложнения структуры теста и увеличения числа заданий в тесте потребность в определении их сложности возрастает. Это особенно важно при сравнении результатов учащихся, решавших разные тесты, или в тех случаях, когда времени на решение всех заданий теста недостаточно.

Так как тесты часто проводят по завершении части или всего курса обучения, вполне естественно определить трудность (сложность, вес) задания как время, необходимое для подготовки учащегося к его выполнению. Если учащийся самостоятельно выбирает последовательность изучаемых разделов (с учётом структуры курса и имеющегося у него времени), то за вес задания можно принять некоторое усреднённое время на подготовку к его решению. Такой способ нахождения веса характерен для теории игр. Таким образом, проблема назначения веса задания педагогического теста сводится к проблеме построения дележа в некоторой кооперативной игре.

Известны две группы методов назначения веса заданий теста: экспертный и частотный. В первой группе эксперт (преподаватель, проводящий тестирование) назначает вес заданий теста, основываясь на собственном опыте. Согласно второму методы определения веса назначаются по результатам пробного тестирования по правилу: чем больше тестируемых решило задание, тем меньший вес ему назначается. Однако применить эти методы сложно при малом опыте работы с курсом.

В 2010–2011 гг. в ПГУПС проводилось тестирование студентов по курсу «Интегральное исчисление». Учащиеся выполняли 24 задания по 9 разделам курса. Обработка итогов тестирования показала хорошую устойчивость частоты решения простых задач и плохую – сложных задач. Показано, что первоначально назначенный (экспертный) вес заданий плохо согласуется с частотным весом, хотя последний неплохо согласуется с весом, полученным из решения соответствующей кооперативной игры. Отдельный интерес представляет вес Шепли для редких заданий, так как там частотный вес оказывается крайне неустойчивым.

Игры с упорядоченным множеством игроков

Определение 1. *Кооперативная игра* [5] $\Gamma = \langle I, v \rangle$ определяется множеством игроков $I = \{1, 2, \ldots, n\}$ и *характеристической функцией игры* v, ставящей в соответствие каждому подмножеству игроков $K \subseteq I$ тот выигрыш, который они могут получить, объединившись в коалицию *K*.

Определение 2. Дележом кооперативной игры $\Gamma = \langle I, v \rangle$ называется неотрицательный вектор $x = (x_1, x_2, \ldots, x_n)$, удовлетворяющий условию коллективной рациональности: $x(I) = \sum_{i \in I} x_i = v(I)$.

Определение 3. Для игрока $i \in I$ в игре $\Gamma = \langle I, v \rangle$ величина $\Delta_i(K) = v(K \cup \{i\}) - v(K)$ называется *вкладом игрока i в коалицию K*, или приращением выигрыша коалиции *K* за счет игрока *i*.

Определение 4. *Вектором Шепли* кооперативной игры $\Gamma = \langle I, v \rangle$ называется дележ $Sh(v) = (Sh_1(v), Sh_2(v), \ldots, Sh_n(v))$, компоненты которого находятся по следующим формулам:

$$Sh_i(v) = \sum_{K \subseteq I} \frac{(n-k-1)!\,k!}{n!} (v(K \cup \{i\}) - v(K)),$$

$$k = |K|; \; n = |I|, \; i \in I. \qquad (1)$$

Хорошо известна следующая вероятностная интерпретация вектора Шепли $Sh(v)$. Предположим, что игроки из множества *I* последовательно появляются на некотором месте встречи, и каждый вновь прибывший игрок $i \in I$ получает величину его вклада $\Delta_i(K)$ в создавшуюся до него коалицию *K*, тогда компонента $Sh_i(v)$ равна математическому ожиданию выигрыша игрока *i* при равновероятности всех порядков появления игроков на общем месте встречи.

Пусть $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ – множество разделов курса обучения, по которым составлен некоторый тест. Мы предполагаем, что курс *A* структурирован и его структура определяется двумя функциями $S(\alpha)$, a_α. Первая из них $S(\alpha)$ определяет множество разделов курса, непосредственно предшествующих разделу α. Вторая a_α указывает время, необходимое для освоения раздела α, если учащийся освоил все разделы из множества $S(\alpha)$.

Определение 5. Будем говорить, что раздел β курса *A* следует за разделом α, или раздел α предшествует разделу β, и записывать $\alpha < \beta$, если найдутся такие разделы $\alpha = \gamma_1, \gamma_2, \ldots, \gamma_k = \beta$ курса *A*, для которых выполняются следующие утверждения: $\alpha \in S(\gamma_2), \gamma_2 \in S(\gamma_3), \ldots, \gamma_{k-1} \in S(\beta)$.

Введённое отношение следования (предшествования) обладает свойством транзитивности, т.е. для любых трёх разделов $\alpha, \beta, \gamma \in A$ верно следующее утверждение: $(\alpha < \beta)$ & $(\beta < \gamma) \Rightarrow (\alpha < \gamma)$. В дальнейшем мы предполагаем, что функция непосредственного предшествования $S(\alpha)$ такова, что порождённое ей отношение следования антирефлексивно на множестве разделов *A*.

Обозначим через $V(\alpha) = \{\beta \in A \mid \beta < \alpha\}$ и $\bar{V}(\alpha) = \{\beta \in A \mid \alpha < \beta\}$ – множества из разделов курса, предшествующих разделу α, и разделов, следующих за разделом α, соответственно, а через $V[K] = V(K) \cup K$.

Пример (сумма знаний [1, 2]). Пусть $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ – частично упорядоченное множество игроков, порожденное отношением непосредственного предшествования *S*(α), а $a = (a_1, a_2, \ldots, a_n)$ – неотрицательный вектор. Определим характеристическую функцию игры $\Gamma = \langle A, v \rangle$ по формуле

$$v(K) = \sum_{\alpha \in V[K]} a_\alpha . \qquad (2)$$

Таким образом, значение $v(K)$ есть время, необходимое для изучения разделов из множества разделов *K*. Перечислим некоторые свойства характеристической функции $v(K)$.

1. *Персональность.* $v(\varnothing) = 0$.

2. *Монотонность и неотрицательность.* Для любых двух совокупностей разделов K_1 и K_2 верно следующее утверждение:

$$K_1 \subset K_2 \Rightarrow v(K_1) \le v(K_2).$$

3. *Ориентация на продвинутые разделы.* Значение характеристической функции на множестве разделов *K* совпадает с её значением на множестве максимальных разделов множества *K*, т.е. $v(K) = v(\max K)$. Иными словами, время, затраченное на изучение разделов из множества *K*, совпадает со временем, затраченным на изучение наиболее «продвинутых» разделов из множества *K*.

Теорема. В кооперативной игре $\Gamma = \langle A, v \rangle$ с частично упорядоченным множеством игроков *A* и характеристической функцией (1) компоненты вектора Шепли могут быть найдены по формуле

$$Sh_\alpha(v) = \sum_{\alpha_j \in V[\alpha]} \frac{a_j}{\left|\bar{V}[\alpha_j]\right|}, \qquad (3)$$

где α – имя игрока; $\left|\bar{V}[\alpha]\right|$ – число элементов множества $\bar{V}[\alpha] = \bar{V}(\beta) \cup \{\beta\}$.

Теорема была сформулирована в работе [2], доказательство ее для древовидного порядка на множестве игроков можно найти в работе [3].

Вес задач педагогического теста

Предположим, что по окончании учебного курса $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ учащийся выполняет проверочный тест, составленный таким образом, что из каждого раздела в тест входит ровно одно задание, поэтому для заданий теста мы будем использовать те же обозначения, что и для соответствующих разделов курса *A*. Если тестируемый выполнил задания совокупности *K* теста $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, то *v*(*K*) можно интерпретировать как уровень его подтверждённых знаний или уровень его подготовки к тесту *A*.

К сожалению, такая оценка уровня знаний крайне неустойчива. Во-первых, выполнение хотя бы одного задания высокого уровня обеспечивает тестируемому высокий балл. Однако такое решение может появиться или случайно, или как результат несанкционированных консультаций. Во-вторых, тестируемый может неправильно оценить свои силы и затратить всё отведённое время на выполнение наиболее сложного задания. В этом случае возможная неудача (недостаток времени, нервозная обстановка, арифметическая ошибка) приведёт к тому, что итоговый балл тестируемого будет сильно занижен. В-третьих, балл тестируемого мало зависит от числа выполненных им заданий, что затрудняет дальнейшее сравнение учащихся.

На практике давно и успешно применяется аддитивная оценка уровня знаний тестируемого:

$$w(K) = \sum_{\alpha \in K} w_\alpha , \qquad (4)$$

где K – множество заданий теста $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, выполненных тестируемым; w_α – вес соответствующих заданий, $\alpha \in A$. В дальнейшем функцию $w(K)$ мы будем называть весовой функцией теста (изучаемого курса). Хорошо известны две группы методов назначения весов заданий теста.

Экспертный. Обычно экспертом является составитель заданий, и он назначает вес, руководствуясь собственным опытом. Возможны и более сложные процедуры назначения веса заданий с участием нескольких экспертов.

Частотный. Вес заданий назначают лишь после проведения теста, причём задание получает больший вес, если оно решено меньшим числом тестируемых. Для расчёта веса используют различные математические модели, в том числе Раша, Бирнбаума и др. [3].

Недостатки обоих методов очевидны. Первый субъективен и мало применим к построению тестов по новым курсам. К недостаткам второго следует отнести невозможность указания веса задания до проведения теста и появление большого веса у простых, но непопулярных заданий.

Перечислим свойства веса Шепли $w = \mathrm{sh}(v)$ и покажем, что он является разумным весом заданий теста v.

Пусть K – совокупность выполненных заданий теста A или совокупность соответствующих разделов курса A. Тогда оценка уровня знаний $w(K)$ близка к истинному уровню знаний $v(K)$ (является оценкой наименьших квадратов с соответствующими коэффициентами).

Весовая функция $w(K)$ аддитивна, положительна и строго монотонна (если время на изучение начальных разделов положительно), т. е.

$$w(K_1 \cup K_2) = w(K_1) + w(K_2) \quad \text{при } K_1 \cap K_2 = \varnothing;$$

$$K_1 \subset K_2 \;\Rightarrow\; w(K_1) < w(K_2) \quad \text{при } K_1,\ K_2 \subseteq A.$$

Вес (балл) Шепли w_α измеряется в часах, связан со временем подготовки к выполнению заданию α курса A и его структурой. Вес w_α задания α – математическое ожидание времени обучения этому заданию при равновероятности всех программ обучения, или доля времени, затраченного строго на изучение раздела α.

Вектор веса w находится в дополнительном *c*-ядре $C^*(v)$ и может быть легко построен по любому курсу A, если известны его структура и функция a_α.

Примеры расчета веса заданий

Пример 1. Предположим, что изученный курс состоит из трех разделов $A = \{\alpha_1, \alpha_2, \alpha_3\}$. Логическая структура курса A описывается функцией непосредственного предшествования $S(\alpha)$, задаваемой равенствами

$$S(\alpha_1) = S(\alpha_2) = \varnothing;\; S(\alpha_2) = \{\alpha_1, \alpha_2\}.$$

Таким образом, разделы α_1 и α_2 являются начальными, а освоение раздела α_3 возможно лишь после освоения двух предыдущих разделов. Обозначим через a_1, a_2 время, необходимое для освоения соответствующих разделов, а через a_3 – время, необходимое для освоения третьего раздела, если предыдущие два раздела уже освоены.

Согласно теореме, вес Шепли соответствующих разделов можно найти по формулам $w_1 = a_1/2$; $w_2 = a_2/2$; $w_3 = a_1/2 + a_2/2 + a_3$.

Заметим, что вес 3-го раздела оказывается больше веса 1-го и 2-го разделов даже в том случае, когда время a_3, затраченное на освоение раздела α_3, равно нулю.

Пример 2. Предположим теперь, что тест, составленный по курсу A, содержит k_1 заданий первого раздела, k_2 заданий второго раздела и k_3 заданий третьего раздела. Составим кооперативную игру «Сумма знаний». В этой игре будет $k_1 + k_2 + k_3$ игроков. Внутри каждого раздела задания можно считать линейно упорядоченными, причем время освоения каждого нового задания считаем равным нулю. Следовательно, вес заданий каждого раздела оказывается равным. Запишем формулы веса для заданий трех разделов: $w_1 = a_1/(k_1 + k_3)$; $w_2 = w_2 = a_2/(k_2 + k_3)$; $w_3 = a_1/(k_1 + k_3) + a_2/(k_2 + k_3) + a_3/k_3$.

Пример 3. На кафедре «Математика и моделирование» в различные годы в рамках курса «Математический анализ» проводилось тестирование студентов по теме «Интегральное исчисление». В 2012, 2013 гг. в тестировании участвовало 120 и 87 человек, соответственно. Задания теста были составлены по 9 разделам, названия которых перечислены в таблице. Обозначим через α_i название i-го раздела проверяемой темы, а через k_i – число заданий теста соответствующего раздела, $i = \overline{1,9}$. Всего тест содержал $\sum_{i=1}^{9} k_i = 24$ заданий различной трудности. Экспертный вес заданий $w_i^{Эк}$ был указан в тесте (см. таблицу), и они должны были помочь тестируемым выбирать задания в соответствии с их уровнем подготовки. Максимально возможная сумма баллов

Баллы заданий теста

№ (i)	Проверяемые разделы темы «Интегрирование» (α_i)	Число заданий раздела в тесте (k_i)	Время освоения раздела (a_i)	Экспертные баллы ($w_i^{Эк}$)	Частотные баллы		Баллы Шепли (w_i^{Sh})
					2012 г. (w_i^{12})	2013 г. (w_i^{13})	
1	Таблица интегралов	10	3	1,0	0,5	1,1	0,5
2	Следствия из таблицы	4	5	3,0	1,2	2,1	2,2
3	Определенный интеграл	1	1	3,0	2,2	4,2	4,6
4	Интегрирование по частям	3	2	5,0	2,0	3,2	3,3
5	Разложение дроби	2	2	3,0	0,9	1,7	1,5
6	Коэффициенты разложения	1	1	3,0	1,8	3,3	2,8
7	Интеграл от дроби	1	1	6,0	2,5	3,9	7,9
8	Двойной интеграл	1	2	10,0	6,4	19,6	12,1
9	Интеграл от тригонометрического выражения	1	3	10,0	44,5	11,8	20,4
	Итого	24	20	75	75	75	75

$\sum_{i=1}^{9} k_i w_i^{\text{Эк}} = 75$, причем сложность заданий, как правило, возрастала при увеличении номера задания.

Для каждого года частота выполнения заданий теста рассчитывалась по формулам $p_i = N_i/(k_i N)$, где N_i – число заданий i-го раздела, выполненных тестируемыми; k_i – число заданий i-го раздела, предложенных в тесте; N – общее число тестируемых в рассматриваемом году. Из таблицы видно, что значения частоты выполнения заданий для двух лет близки (кроме наиболее сложной задачи раздела 9). Частотные баллы строились по следующим формулам:

$$w_i = \frac{k}{p_i};\ k = 75 \cdot \left(\sum_{i=1}^{9} \frac{k_i}{p_i}\right)^{-1},$$

в которых k – нормирующий множитель, обеспечивающий максимальную общую сумму баллов, равную 75. В таблице можно заметить, что полученные частотные баллы значительно отличаются как от экспертных баллов, так и друг от друга в различные годы тестирования.

Тема «Интегрирование» изучалась в соответствии с учебным планом и указанными в нем взаимосвязями разделов. Отношение непосредственного предшествования разделов может быть задано функцией $S(\alpha)$, ее значения:

$$S(\alpha_1) = S(\alpha_5) = \varnothing;\ S(\alpha_2) = \{\alpha_1\};$$
$$S(\alpha_6) = \{\alpha_5\};\ S(\alpha_4) = \{\alpha_2\};\ S(\alpha_3) = \{\alpha_4\};$$
$$S(\alpha_8) = \{\alpha_3\};\ S(\alpha_7) = \{\alpha_2, \alpha_6\};$$
$$S(\alpha_9) = \{\alpha_3, \alpha_7\}.$$

Согласно определению 5, функции $S(\alpha)$ на множестве разделов A порождают отношение следования. Таким образом, 3-й и 7-й разделы оказались начальными, а 8-й и 9-й указаны как наиболее сложные.

Числа $a_i \geq 0$, $i = \overline{1,9}$ (см. таблицу) определяют время, необходимое для изучения соответствующих разделов, при условии, что все предыдущие разделы уже изучены. Их значения указаны в таблице. Таким образом, время на изучение всех разделов изучаемой темы равно $a_1 + a_2 + \cdots + a_9 = 20$ (час).

Так как по некоторым разделам тест содержал несколько заданий, для построения веса Шепли необходимо определить кооперативную игру с числом участников, равным 24. Время подготовки участников к выполнению заданий одного и того же раздела теста мы считаем равным времени освоения этого раздела. Так как общая сумма веса Шепли равна общему времени освоения темы, т. е. 20, то баллы Шепли будут в $k = 75/20 = 3,25$ раз больше соответствующего веса, что сделано для удобства сравнения с ранее рассчитанными баллами.

Окончательно баллы Шепли были рассчитаны по формулам:

$$w_1 = k \cdot a_1/21;\ w_2 = k(a_1/21 + a_2/11);$$
$$w_4 = k(a_1/21 + a_2/11 + a_4/7);\ w_5 = k \cdot a_5/5;$$
$$w_3 = k(a_1/21 + a_2/11 + a_3/3 + a_4/7);$$
$$w_6 = k \cdot (a_5/5 + a_6/3);$$
$$w_7 = k(a_1/21 + a_2/11 + a_4/7 + a_5/5 + a_6/3 + a_7/2);$$
$$w_8 = k(a_1/21 + a_2/11 + a_3/3 + a_4/7 + a_8);$$
$$w_9 = k(a_1/21 + a_2/11 + a_3/3 + a_4/7 + a_5/5 + a_6/3 + a_7/2 + a_9).$$

Их точные значения приведены в таблице.

Из таблицы видно, что баллы Шепли близки к частотным баллам для заданий с малыми номерами, а для последних заданий их значения лежат между частотными баллами различных лет.

Вывод

При проведении итогового или промежуточного контроля баллы Шепли могут быть хорошим ориентиром при назначении баллов заданий теста.

Библиографический список

1. **Теория** статистических решений : учеб. пособие. Ч. 2 / М. М. Луценко. – Санкт-Петербург : Петербургский гос. ун-т путей сообщения, 2012. – 111 с.

2. **Lutsenko, M. M.,** Shadrintseva, N. V. Shapley value in testing // Proceedings of the 2nd int. conf. «Game theory and management application», «Game theory, operations research and applications», Hyderabad, Inst. of public enterprise, India. – Hyderbad, 2012. – P. 47–49.

3. **Введение** в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – Москва, 2000. – 168 с.

4. **О точности** педагогического тестирования / М. М. Луценко, Н. В. Шадринцева // Изв. Петербург. ун-та путей сообщения. – Санкт-Петербург : Петербург. гос. ун-т путей сообщения, 2011. – Вып. 4 (29). – С. 250–258.

5. **Теория** игр : учеб. пособие для университетов / Л. А. Петросян, Н. А. Зенкевич, Е. А. Семина. – Москва : Высш. шк., 1998. – 304 с.