

ЭКОНОМИКА

УДК 378.146

Е.В. Жилина

ИСПОЛЬЗОВАНИЕ БИНАРНОЙ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ
ДЛЯ ОЦЕНКИ КАЧЕСТВА АДАПТИВНОГО ТЕСТА

Приводится обоснование использования метода бинарной логистической регрессии для оценки качества адаптивного теста; приведен пример адаптивного тестирования, рассчитаны вероятности получения правильных ответов тестируемыми на каждом уровне знания.

Ключевые слова: адаптивный тест; оценка качества; логистическая регрессия.

Адаптивный тест – это один из видов проверки знаний учащихся, предусматривающий изменение последовательности выдачи вопросов в процессе прохождения теста, учитывающий ответы тестируемого на предыдущие вопросы для определения трудности последующих вопросов [1]. Адаптивность сочетается с принципом «алгоритма-цепочки»: предъявление заданий с систематическим изменением уровня трудности.

Адаптивный тест как система обладает составом, целостностью и структурой. Тест состоит из заданий, правил их применения, оценок за выполнение каждого задания и рекомендаций по интерпретации тестовых результатов. Система означает, что в тесте собраны такие задания, которые обладают системообразующими свойствами. Хотя любой тест состоит из тестовых заданий, последние представляют не совокупность произвольно объединенных заданий, а именно систему. Время выделяют в качестве другого системообразующего фактора. Действительно, одно из соображений, положенных в основу создания адаптивных тестов, – иметь инструмент быстрого и относительно точного оценивания большого числа испытуемых. Требование экономии времени становится естественным в массовых процессах, каковым и стало образование. Одно из актуальных направлений современной организации тестового контроля – это индивидуализация контроля, приводящая к значительной экономии времени тестирования. От времени тестирования существенно зависит качество результатов. Каждый адаптивный тест имеет оптимальное время тестирования, необходимое для получения точной оценки тестируемого, уменьшение или превышение которого снижает качественные показатели теста.

Эффективным можно назвать тест, который лучше, чем другие тесты, измеряет знания студентов интересующего уровня подготовленности, с меньшим числом заданий, качественнее, быстрее, дешевле, и все это – по возможности в комплексе [2]. С понятием «эффективность» сопряжено и близкое к нему по содержанию понятие «оптимальность». Последнее трактуется как наилучшее из возможных вариантов, с точки зрения удовлетворения нескольким критериям, взятым поочередно или вместе. В определении эффективности теста учитываются два ключевых элемента: число заданий теста и уровень подготовленности студентов. Если из какого-либо теста с большим числом заданий сделать оптимальный выбор меньшего числа, то может образоваться система, не уступающая заметно по своим свойствам тесту со сравнительно большим числом заданий.

Тест с меньшим числом заданий в таком случае можно называть сравнительно более эффективным.

Помимо этого, эффективность теста можно оценить с точки зрения соответствия уровня его трудности уровню подготовленности тестируемых в данный момент студентов. Эту оценку в литературе нередко относят к валидности, имея в виду идею валидности теста, так сказать, по уровню. Легко понять практическую бесполезность того, чтобы давать слабым студентам трудные задания; большинство студентов, вероятнее всего, не сумеют правильно на них ответить. Так же обстоит дело и с легким тестом: его бесполезно (неэффективно) давать знающим студентам, потому что и здесь высока вероятность теперь уже правильных ответов, и потому практически все испытуемые получают по тесту одинаково высокий балл. И в том и в другом случае испытуемые не будут различаться между собой. Измерение, таким образом, не состоится по причине несоответствия уровня трудности теста уровню подготовленности. Из этих соображений легко вывести, что самый эффективный тест – это тест, точно соответствующий по трудности заданий уровню подготовленности испытуемых.

Эффективность теста зависит также и от принципа подбора заданий. Если подбирать задания для измерения на всем диапазоне изменения трудности, то снижается точность измерения на отдельном участке. И наоборот, если стремиться точно измерить знания испытуемых, например среднего уровня подготовленности, то это потребует иметь больше заданий именно данного уровня трудности. Поэтому тест не может быть эффективным вообще, на всем диапазоне подготовленности студентов. Он может быть более эффективен на одном уровне знаний и менее – на другом. Именно этот смысл вкладывается в понятие дифференциальной эффективности адаптивного теста.

С помощью метода бинарной логистической регрессии можно исследовать зависимость дихотомических переменных от независимых переменных, имеющих любой вид шкалы. Как правило, в случае с дихотомическими переменными речь идет о некотором событии, которое может произойти или не произойти; бинарная логистическая регрессия в таком случае рассчитывает вероятность наступления события в зависимости от значений независимых переменных [3].

Как известно, все регрессионные модели могут быть записаны в виде формулы

$$Y = F(x_1, x_2, \dots, x_n). \quad (1)$$

Например, в множественной линейной регрессии предполагается, что зависимая переменная является линейной функцией независимых переменных, т.е.

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n, \quad (2)$$

где x_1 – значения независимых переменных, b_1 – коэффициенты, расчёт которых является задачей бинарной логистической регрессии, a – некоторая константа.

Можно ли ее использовать для задачи оценки вероятности исхода события? Да, можно, вычислив стандартные коэффициенты регрессии. Например, если рассматривается исход по ответу на тестовое задание, задается переменная y со значениями 1 и 0, где 1 означает, что тестируемый правильно ответил на вопрос, а 0 – что неправильно. Однако здесь возникает проблема: множественная регрессия не «знает», что переменная отклика бинарна по своей природе. Это неизбежно приведет к модели с предсказываемыми значениями большими 1 и меньшими 0. Но такие значения не допустимы для первоначальной задачи. Таким образом, множественная регрессия просто игнорирует ограничения на диапазон значений для y .

Для решения проблемы задача регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной мы предсказываем непрерывную переменную со значениями на отрезке $[0,1]$ при любых значениях независимых переменных. Это достигается применением следующего регрессионного уравнения (логит-преобразование):

$$P = \frac{1}{1 + e^{-y}}, \quad (3)$$

где P – вероятность того, что произойдет интересное событие; e – основание натуральных логарифмов 2,71...; y – стандартное уравнение регрессии [4].

Если для P получится значение меньше 0,5, то можно предположить, что событие не наступит; в противном случае предполагается наступление события.

Поясним необходимость преобразования. Предположим, что мы рассуждаем о нашей зависимой переменной в терминах основной вероятности P , лежащей между 0 и 1. Тогда преобразуем эту вероятность P :

$$P' = \log_e(P / (1 - P)). \quad (4)$$

Это преобразование обычно называют логистическим или логит-преобразованием. Теоретически P' может принимать любое значение. Поскольку логистическое преобразование решает проблему об ограничении на 0–1 границы для первоначальной зависимой переменной (вероятности), то эти преобразованные значения можно использовать в обычном линейном регрессионном уравнении. А именно: если произвести логистическое преобразование обеих частей описанного выше уравнения, мы получим стандартную модель линейной регрессии [4].

Существует несколько способов нахождения коэффициентов логистической регрессии. На практике часто используют метод максимального правдоподобия.

Нами предлагается использовать метод бинарной логистической регрессии для оценки вероятностей получения правильного ответа на задание по каждому уровню знания как один из критериев качества адаптивного теста, уменьшающий время его прохождения, число заданий и увеличивающий точность оценки.

Предположим, что испытуемый начал тестирование с вопроса, соответствующего уровню 3, шаг контрольной точки равен 5, максимальное количество шагов тестирования равно 6, на первом шаге адаптации были получены следующие результаты (рис. 1, шаг 1; «+» отмечен правильный ответ, «-» – неправильный) (на основе метода, предложенного в (5)).

уровень сложности(L) \ вопрос	вопрос																				R _{Li}		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20			
7																						0,00	
6					-																	0,00	
5				+		-															-	0,25	
4			+				+		-				-		-					+		-	0,43
3		+								-			+		+		-			+			0,67
2											+						+						1,00
1																							0,00
		шаг 1				шаг 2				шаг 3				шаг 4									

Рис. 1. Пример адаптивного теста

«Промежуточные уровни знания» используются во избежание случайности ответов (как положительных, так и отрицательных). Поэтому под термином «промежуточные уровни знания» понимаем логическое распределение уточняющих вопросов на оценки 3, 4, 5 (уровни знания соответственно равны 3, 5, 7).

Используя программный продукт STATISTICA 6.0, были обработаны результаты тестирования каждого шага адаптации. Исходные данные приведены на рис. 2.

Гипотеза $H(0)$ предполагает, что тестируемый ответит на вопрос правильно (1).

На первом шаге адаптации были получены следующие результаты логистической регрессии (см. рис. 3).

Далее, согласно формуле 3, вычисляем вероятности (P_i) того, что тестируемый на данном шаге адаптации ответит правильно на вопрос соответствующего уровня сложности (см. рис. 4). P_0 – начальное значение – принимаем равным 0,5 на каждом уровне сложности. Рассчитываем коэффициент изменения вероятности

$$\Delta P : \Delta P = P_{i+1} - P_i. \quad (5)$$

Согласно алгоритму, предложенному в [5], шагов адаптации должно быть не менее двух, поэтому тестируемому предлагается продолжить тест далее (рис. 1, 2, шаг 2). На втором и третьем шаге адаптации были получены следующие результаты логистической регрессии

сии (рис. 5, а, б) и рассчитаны вероятности ($P_2, P_3, \Delta P$) (рис. 4).

На данных этапах тестирования р-уровень более 5%, следовательно, модели незначимы, поэтому продолжаем тестирование. Далее приведены результаты четвертого шага адаптации (см. рис. 6).

Вычисляем вероятности ($P_4, \Delta P$) (см. рис. 7).

Анализируем полученные результаты: р-уровень менее 5%, следовательно, модель значима; значение статистики хи-квадрат для разницы между текущей моделью и моделью, содержащей лишь свободный член, высоко значимо.

	номер вопроса	Ответ	Ур_сложности
шаг 1	1	1	3
	2	1	4
	3	1	5
	4	0	6
	5	0	5
шаг 2	6	1	4
	7	0	5
	8	0	4
	9	0	3
	10	1	2
шаг 3	11	1	3
	12	0	4
	13	1	3
	14	0	4
	15	0	3
шаг 4	16	1	2
	17	1	3
	18	1	4
	19	0	5
	20	0	4

Рис. 2. Исходные данные для моделирования логистической регрессии

		Model: Logistic regression (logit) N of 0's:2 1's:3 (Spreadsheet2)			
		Dep. var: Ответ Loss: Max likelihood			
		Final loss: 1,386319092 Chi?(1)=3,9575 p=,04667			
N=5	Const. B0	Ур_сложности			
Estimate	92,90471	-18,5789			
Odds ratio (unit ch)		0,0000			
Odds ratio (range)		0,0000			

Рис. 3. Результаты логистической регрессии 1-го шага адаптации теста

ур_сложности	P0	P1	ΔP	P2	ΔP	P3	ΔP
2	0,5	1	0,5	0,874384	-0,125616	0,822591111	-0,0517931
3	0,5	1	0,5	0,735649	-0,264351	0,650010069	-0,085639
4	0,5	1	0,5	0,52664	-0,47336	0,426575933	-0,1000638
5	0,5	0,502482	0,0024824	0,307857	-0,194625	0,229567712	-0,0782896
6	0,5	9,12E-09	-0,5	0,150975	0,1509753	0,106626531	-0,0443488
7	0,5	8,24E-17	-0,5	0,066373	0,066373	0,045625432	-0,0207475
p		0,04667		0,14699		0,1071	
Chi-square		3,9575		2,1034		2,5967	

Рис. 4. Рассчитанные вероятности наступления события «1» на 1–3-м шагах адаптации

		Model: Logistic regression (logit) N of 0's:2 1's:3 (Spreadsheet2)	
		Dep. var: Ответ Loss: Max likelihood	
		Final loss: 9,065524176 Chi?(1)=2,5967	
N=15	Const. B0	Ур_сложности	
Estimate	3,37413	-0,917719	
Odds ratio (unit ch)	29,19902	0,399429	
Odds ratio (range)		0,025454	

а

		Model: Logistic regression (logit) N of 0's:2 1's:3 (Spreadsheet2)	
		Dep. var: Ответ Loss: Max likelihood	
		Final loss: 5,879795618 Chi?(1)=2,1034	
N=10	Const. B0	Ур_сложности	
Estimate	3,78547	-0,919622	
Odds ratio (unit ch)	44,05653	0,398670	
Odds ratio (range)		0,025261	

б

Рис. 5. Результаты логистической регрессии 2-го (а) и 3-го (б) шагов адаптации теста

		Model: Logistic regression (logit) N of 0's:10 1's:10 (4. sta)		
		Dep. var: Ответ Loss: Max likelihood		
		Final loss: 11,224787155 Chi?(1)=5,2763 p=.02162		
N=20		Const. B0	Ур_сложности	
Estimate	4,52558		-1,19415	
Odds ratio (unit ch)	92,34928		0,30296	
Odds ratio (range)			0,00842	

Рис. 6. Результаты логистической регрессии четвертого шага адаптации теста

ур_сложности	P4	ΔP
2	0,893857171	0,07126606
3	0,719151396	0,06914133
4	0,437762443	0,01118651
5	0,19142888	-0,0381388
6	0,067153695	-0,0394728
7	0,021420351	-0,0242051
p		
Chi-square	0,0216239	5,276313
-2*log(Likelihood)		
	for this model = 22,44957 intercept only: 27,72589	

Рис. 7. Рассчитанные вероятности наступления события «1» на 4-м шаге адаптации

Одним из критериев оценки качества логистической регрессии является отношение несогласия (Odds ratio) (рис. 8), полученное методом классификации наблюдений.

Отношение несогласия вычисляется как отношение произведения чисел правильно расклассифицированных наблюдений к произведению чисел неправильно расклассифицированных. Отношение несогласия больше 1 показывает, что построенная классификация лучше, чем если бы мы просто провели классификацию наугад.

Нами предлагается интерпретировать результаты классификации наблюдений для установления количества угаданных ответов и количества случайных ошибок (рис. 8).

В рассматриваемом нами примере адаптивного теста четвертый шаг адаптации является заключительным.

	Pred.0	Pred.1	Percent
0,000000	8	2	80,00000
1,000000	4	6	60,00000

Odds ratio: 6,0000

угадывание	2
случайная ошибка	4

Рис. 8. Отношение несогласия

Ур_знания	Оценка
7	
6	5
5	
4	4
3	
2	3
1	2

Рис. 9. Шкала перевода уровня знания (уровня сложности) в оценку

ЛИТЕРАТУРА

1. *Тестология* в России: модели и методы тестологии. Режим доступа: <http://www.testor.ru>
2. Тягунова Т.Н. *Философия и концепция компьютерного тестирования*. М.: МГУП, 2003. 246 с.
3. *Иллюстрированный самоучитель по SPSS*. Режим доступа: <http://lib.qrz.ru/node/11329>
4. Паклин Н. Логистическая регрессия и ROC-анализ – математический аппарат. Режим доступа: <http://www.basegroup.ru/library/analysis/regression/logistic>
5. Жилина Е.В. *Линейная модель адаптивного тестирования // Проблемы создания и использования информационных систем и технологий: Материалы III межрегион. конф. Ростов н/Д: РИНХ, 2009.*

Статья представлена научной редакцией «Экономика» 1 марта 2010 г.