



ПРИМЕНЕНИЕ ПОРЯДКОВЫХ СТАТИСТИК ПРИ АНАЛИЗЕ ВЗАИМОСВЯЗЕЙ МЕЖДУ ЗАДАНИЯМИ ТЕСТОВ

ШЕРГИН В.Л.

Проводится анализ устойчивости числовых характеристик по отношению к выбросам и неоднородности выборки. Предлагается новая робастная оценка степени статистической взаимосвязи случайных величин. Численный эксперимент показал, что при оценивании статистической взаимосвязи между временными характеристиками выполнения тестовых заданий предложенная оценка является устойчивой и более чётко отражает взаимосвязи, чем традиционные коэффициенты Пирсона и Кендалла. Предложенная оценка может быть использована в системах автоматизации анализа качества тестовых заданий.

Введение

В настоящее время тестовые формы контроля знаний студентов считаются одним из наиболее перспективных средств обеспечения оперативности и объективности контроля знаний. При этом, как отмечают специалисты [1, 2], если на первых этапах внедрения тестовых технологий в учебный процесс основной задачей являлось создание банков тестовых заданий по дисциплинам, а также решение проблем программного и технического обеспечения процесса тестирования, то на современном этапе на первый план выходит разработка и реализация эффективных инструментов анализа тестовых заданий.

Важнейшим инструментальным средством автоматизации анализа качества тестов является статистическая обработка результатов тестирования. При этом несмотря на компьютеризацию процесса тестирования и подведения его итогов, статистический анализ результатов проводится достаточно редко и поверхностно, зачастую сводится лишь к выявлению чрезмерно простых и чрезмерно сложных вопросов теста.

Статистические методы достаточно давно и широко используются при анализе результатов тестирования «по строкам», т.е. в целях оценивания или рейтингования студентов [3, 4]. В то же время их применение для оценивания качества самих тестов сдерживается как по организационным причинам (увеличение временных и интеллектуальных затрат на формирование конечного варианта теста, что воспринимается как нецелесообразный шаг), так и в силу несовершенства

(или неправильного применения) самих статистических методов.

1. Анализ проблемной области и постановка задачи

Известно, что качество теста во многом определяется качеством начального тестового материала. При этом основным критерием оценки качества теста является полнота отображения в тестовом материале содержания дисциплины [1]. Очевидно, что этот критерий является недостаточным, поскольку он никак не отражает такие аспекты качества вопросов, как однозначность формулировки, наличие скрытых подсказок, взаимосвязь между вопросами.

При этом следует отметить, что анализ результатов тестирования является важным инструментом поддержания обратной связи между студентами и преподавателями. Только при ее наличии преподаватель сможет скорректировать изложение материала, выделить потенциально трудные темы в целях их более подробного рассмотрения, учесть степень восприятия материала при изложении обеспечиваемых дисциплин. Эта проблема является особенно актуальной в условиях, когда обратная связь с преподавателем затруднена, либо текущий контроль знаний отложен по времени, что характерно для заочной и дистанционной форм обучения [2, 4].

Основным методом анализа качества тестов в настоящее время есть экспертный. Очевидными недостатками этого метода являются сложность формализации и, следовательно, автоматизации анализа, высокая трудоёмкость проведения анализа, а также в ряде случаев сложности, связанные с организацией экспертизы и подбором экспертов [3].

Основным препятствием на пути использования статистических методов анализа качества теста является приверженность к «классическим» статистическим методам. Эти методы и соответствующие параметрические оценки основаны на гипотезах об однородности выборки, гауссовых распределениях и линейных взаимосвязях. Но поскольку для выборок, представляющих собой результаты тестирования, эти гипотезы не соблюдаются, то и получаемые результаты такого статистического анализа имеют низкую надёжность и низкую дифференцирующую способность.

Целью исследования является разработка новых статистических методов, обеспечивающих получение устойчивых оценок параметров неоднородных выборок. Эта задача представляет большой интерес. В то же время робастное оценивание степени статистических взаимосвязей между заданиями тестов позволит автоматизировать и усовершенствовать процесс анализа качества тестовых заданий, что является актуальной прикладной задачей.

Как известно [5, 6], наибольшей устойчивостью по отношению к выбросам и другим неоднородностям выборки обладают оценки, основанные на порядковых статистиках. По этой причине предлагается для

оценки качества тестов использовать непараметрические методы анализа, основанные на порядковых статистиках.

2. Устойчивое оценивание параметров выборки

В настоящее время в прикладной статистике, как отмечают специалисты [7], господствующее положение занимают «классические» методы оценивания числовых характеристик выборки, основанные на двух базовых гипотезах: об однородности выборки и о гауссовости закона распределения. Гипотеза об однородности означает, что все элементы выборки получены из одной и той же генеральной совокупности и, соответственно, следуют одному и тому же закону распределения. Гипотеза о гауссовости усиливает гипотезу об однородности и постулирует, что этот закон распределения является нормальным (гауссовым). Именно на этих двух гипотезах основана большая часть методов дисперсионного, корреляционного, регрессионного, а также факторного и дискриминантного анализов. В рамках этих гипотез оценками факторов сдвига, масштаба и силы линейной взаимосвязи являются соответственно среднеарифметическое значение (m), среднеквадратичное отклонение (s) и коэффициент корреляции Пирсона r_{xy} :

$$m_x = \frac{1}{n} \sum x_i, \quad s_x = \sqrt{\frac{1}{n} \sum (x_i - m)^2}, \quad (1)$$

$$r_{xy} = \frac{\frac{1}{n} \sum ((x_i - m_x)(y_i - m_y))}{\sqrt{\frac{1}{n} \sum (x_i - m_x)^2} \cdot \sqrt{\frac{1}{n} \sum (y_i - m_y)^2}}. \quad (2)$$

Очевидно, что выборки, содержащие реальные данные, далеко не всегда следуют этим гипотезам, а даже если и следуют, то нет оснований принимать этот факт а priori. Но что гораздо хуже, прикладные статистики редко отдают себе отчет в том, что при нарушении гипотез о нормальности и однородности оценки (1)-(2) не только не являются эффективными, но и могут быть не состоятельными.

Именно это обстоятельство зачастую и является причиной низкой устойчивости оценок и, как следствие, низкого качества результатов статистического анализа.

Объектом исследования для порядковых статистик служит упорядоченная (отсортированная) выборка. Основным достоинством оценок, основанных на порядковых статистиках, является их устойчивость по отношению к статистическим свойствам выборки и, в частности, к выбросам и к возможной неоднородности выборки [6]. Устойчивость проявляется в том, что сколь угодно большое изменение отдельных элементов выборки приводит к конечному (и обычно незначительному) изменению оценки. При этом вариация оценки зависит не от величины выбросов, а от количества таких элементов. Следствием этого является тот факт, что при неоднородности выборки получаемые оценки всегда характеризуют свойства той её

части, элементы которой составляют большинство (если оно есть), либо части, нейтральной по расположению.

Примерами оценок, основанных на порядковых статистиках, являются медианные оценки: медиана выборки (h) и медиана абсолютных отклонений от медианы (median absolute deviation, MAD):

$$h_x = \text{median}(x_i), \quad i = 1, \dots, n$$

$$\text{MAD}_x = \sqrt{\text{median}((x_i - h_x)^2)} = \text{median}(|x_i - h_x|). \quad (3)$$

Медианные оценки (3) используются в качестве мер сдвига и масштаба вместо традиционных оценок m и s (1) [8]. Нетрудно заметить, что с формальной точки зрения оценки (3) отличаются от оценок (1) только тем, что при их вычислении вместо оператора взятия среднего арифметического $\frac{1}{n} \sum ()$ используется оператор взятия медианы $\text{median}()$.

В настоящей работе предлагается применить тот же подход и при оценивании степени взаимосвязи двух случайных величин, заменив коэффициенты корреляции Пирсона r_{xy} (2) следующими коэффициентами взаимосвязи ρ_{xy} :

$$\rho_{xy} = \frac{\text{median}((x_i - h_x)(y_i - h_y))}{\sqrt{\text{median}((x_i - h_x)^2)} \cdot \sqrt{\text{median}((y_i - h_y)^2)}} = \frac{\text{median}((x_i - h_x)(y_i - h_y))}{\text{MAD}_x \cdot \text{MAD}_y}. \quad (4)$$

Как и другие медианные оценки (3), предложенный коэффициент взаимосвязи мало чувствителен к выбросам, в отличие от коэффициента корреляции (2). Вариация коэффициента (4) в ответ на выбросы является малой и не зависит от величины выбросов. При малых же изменениях элементов выборки, не влияющих на их ранги и ранги произведений, коэффициент (4), в отличие от (2), вообще не изменяется. В этом смысле предложенный коэффициент проявляет сходство с коэффициентами ранговой корреляции (R Спирмена и τ Кендалла [9]).

При этом, однако, коэффициент (4) основан на количественной шкале измерения элементов выборки, а не на порядковой, и отражает не только упорядоченность рангов элементов, но и количественные соотношения между их значениями. Поэтому можно сказать, что предложенный коэффициент связи (4) обладает компромиссными свойствами между коэффициентами ранговой корреляции и коэффициентом Пирсона.

Одним из следствий устойчивости предложенных коэффициентов взаимосвязи (4) по отношению к выбросам и неоднородности выборки является гораздо лучшая обусловленность матрицы этих коэффициентов по сравнению с корреляционной матрицей. Если выборка действительно неоднородна, то элементы

предложенной матрицы будут гораздо ближе к характерным значениям 0, ± 1 , и в то же время количество элементов, близких к ± 1 , в ней будет меньше, чем в обычной корреляционной матрице. Таким образом, можно сказать, что матрица, состоящая из предложенных коэффициентов взаимосвязи (4), будет не только менее чувствительна к шумам и выбросам, но и будет более чётко отражать взаимосвязи между соответствующими объектами.

3. Статистическое оценивание взаимосвязей между вопросами теста

Для анализа качества теста традиционно используются результаты тестирования, представленные в матричной форме и заданные, как правило, в дихотомической шкале. Однако, как отмечают [1, 2], показатели времени выполнения заданий тестов несут гораздо больше информации о качестве теста, чем собственно результаты выполнения заданий.

Одним из методов анализа временных характеристик выполнения теста является корреляционный анализ. Следует отметить, что с точки зрения статистик и задача оценивания качества тестов (т.е. вопросов) существенно более сложная, чем задача оценивания тестируемых (т.е. студентов), что обусловлено анизотропностью выборки. Так, вопросы теста имеют известную сложность. Сложность вопроса задаётся заранее составителем теста, является (по его мнению) объективной и не меняется по годам или академическим группам. С другой стороны, успеваемость студентов тестируемой группы заранее неизвестна и, как показывает практика, может существенно меняться как по годам, так и по группам, потокам, специальностям. На языке математической статистики это означает, что выборка, взятая по студентам, принадлежит генеральной совокупности с известными статистическими свойствами, а выборка, взятая по заданиям теста, – с неизвестными. Именно по этой причине предлагается для оценки степени взаимосвязей между заданиями теста использовать вместо традиционных коэффициентов корреляции (2) непараметрическую оценку (4).

В качестве примера рассмотрим табл.1, показывающую время выполнения студентами теста. В следующих таблицах (табл.2-4) содержатся корреляционная матрица, матрица коэффициентов взаимосвязи (4) и матрица коэффициентов ранговой корреляции τ Кендалла.

Сравнение этих таблиц демонстрирует, что предлагаемые коэффициенты взаимосвязи (4) более чётко отражают зависимости между временами выполнения тестовых заданий, чем коэффициенты Пирсона и Кендалла, показывая три сильных взаимосвязи вместо семи при том, что и сами связи выражены сильнее. Более подробный анализ подтвердил обоснованность такого сокращения, поскольку большие

значения коэффициентов корреляции Пирсона были обусловлены выбросами – аномальными значениями времён выполнения отдельных заданий отдельными студентами.

Таблица 1. Время выполнения теста

		Вопросы							
		v1	v2	v3	v4	v5	v6	v7	v8
Студенты	S1	28	12	20	30	22	34	40	53
	S2	29	13	24	36	25	27	46	54
	S3	30	15	18	34	23	40	44	58
	S4	22	19	24	33	33	26	44	40
	S5	27	10	23	29	20	32	47	33
	S6	29	12	22	20	27	48	42	62
	S7	25	24	20	24	30	37	43	39
	S8	26	20	17	39	29	31	47	59
	S9	25	20	22	30	34	26	38	45
	S10	27	18	24	40	40	33	35	60

Таблица 2. Корреляционная матрица

v1	v2	v3	v4	v5	v6	v7	v8
1							
-0,65	1						
-0,22	-0,23	1					
-0,04	0,14	-0,02	1				
-0,53	0,68	0,32	0,31	1			
0,56	-0,26	-0,33	-0,58	-0,27	1		
0,05	-0,21	-0,27	-0,03	-0,65	-0,05	1	
0,60	-0,14	-0,29	0,28	0,14	0,43	-0,27	1

Таблица 3. Матрица коэффициентов взаимосвязи (4)

	v1	v2	v3	v4	v5	v6	v7	v8
v1	1							
v2	-0,82	1						
v3	0,00	0,00	1					
v4	0,00	0,41	0,21	1				
v5	-0,50	0,79	0,00	0,41	1			
v6	0,15	-0,34	-0,17	-0,33	-0,21	1		
v7	0,00	-0,21	0,08	0,41	-0,12	-0,25	1	
v8	0,48	-0,16	0,00	0,44	-0,13	0,18	0,07	1

Таблица 4. Матрица значений τ Кендалла

v1	v2	v3	v4	v5	v6	v7	v8
1							
-0,42	1						
-0,12	-0,19	1					
0,02	0,14	0,07	1				
-0,41	0,52	0,19	0,27	1			
0,49	-0,23	-0,31	-0,25	-0,22	1		
0,05	-0,12	-0,19	0,14	-0,39	-0,05	1	
0,41	-0,07	-0,14	0,40	0,20	0,31	-0,16	1

Выводы

Проведенный анализ показал, что статистические выборки, содержащие реальные экспериментальные данные, зачастую являются неоднородными. В этом случае для оценивания их параметров необходимо применять методы, устойчивые по отношению к выб-

росам и неоднородности выборки. Таким требованиям удовлетворяют методы, основанные на использовании порядковых статистик.

Научная новизна данной работы состоит в том, что предложен новый метод оценивания степени статистической взаимосвязи между двумя случайными величинами, основанный на порядковых статистиках – коэффициент взаимосвязи (4). Предложенный коэффициент, в отличие от коэффициента корреляции Пирсона, является устойчивым по отношению к выбросам, негауссовости и неоднородности выборки.

Предлагаемые коэффициенты взаимосвязи более чётко отражают зависимости между временами выполнения тестовых заданий, чем коэффициенты Пирсона и Кендалла, показывая три сильных взаимосвязи вместо семи при том, что и сами связи выражены сильнее. Этот результат иллюстрирует практическую значимость работы.

Таким образом, результаты вычислительного эксперимента, проведенного на реальных данных, подтверждают работоспособность и эффективность предложенного метода устойчивого оценивания силы статистических взаимосвязей.

Литература: 1. Ризун Н.О. Эвристический алгоритм совершенствования технологии оценки качества тестовых заданий [Текст] / Н.И.Ризун // Восточно-Европейский журнал передовых технологий. 2010. №3/11 (45). С. 40-48. 2. Белоус Н. Методика определения качества тестовых заданий, оцениваемых по непрерывной шкале [Текст] / Н.Белоус, И. Куцевич, И.Белоус // International Book Series "Information Science and Computing". Kyiv, 2009. P.127-133. 3. Мамонтова М.Ю. Квалиметрический подход к моделированию оценки качества академической подготовки студентов [Текст] / М.Ю. Мамонтова // Успехи современного естествознания. 2007. № 12. С. 145. 4. Шергин В.Л. Использование медианного метода оценки среднего уровня знаний [Текст] / В.Л.Шергин, В.В.Лапина // Управління розвитком. 2007. №7. С. 152-153. 5. Ефимов А.Н. Порядковые статистики – их свойства и приложения [Текст] / А.Н. Ефимов. М.: Знание, 1980. 64с. 6. Дэйвид Г. Порядковые статистики [Текст] / Г. Дэйвид. М.: Наука, 1979. 336с. 7. Орлов А. И. Прикладная статистика [Текст] / А.И. Орлов. М.: Экзамен, 2006. 672с. 8. Шергин В.Л. Робастные методы оценивания числовых характеристик выборки [Текст] / В.Л. Шергин, Т.А. Мирошниченко // Бионика интеллекта. 2010. №3(74). С.90-93. 9. Гайдышев И. Анализ и обработка данных: специальный справочник [Текст] / И. Гайдышев. СПб: Питер, 2001. 784с.

Поступила в редколлегию 17.06.2011

Рецензент: д-р техн. наук, проф. Удовенко С.Г.

Шергин Вадим Леонидович, канд. техн. наук, доцент кафедры искусственного интеллекта ХНУРЭ. Адрес: Украина, 61166, Харьков, пр. Ленина, 14.