

Редюк Виктория Александровна

СИСТЕМЫ АВТОМАТИЧЕСКОГО ТЕСТИРОВАНИЯ ЗНАНИЙ: АЛГОРИТМЫ, ИНСТРУМЕНТЫ, АДЕКВАТНОСТЬ ПОЛУЧЕННЫХ ДАННЫХ

В данной статье рассматривается проблема анализа результатов в системах автоматического тестирования знаний. Автор дает обобщенную характеристику типам тестов, проводит сравнение алгоритмов тестирования, применяемых для автоматических тестов, указывает причины предпочтительности данного вида тестирования остальным. Выделяются и описываются характерные особенности основных алгоритмов, а также проблемы, которые могут возникнуть при использовании каждого из них. Автор обращает внимание, что в настоящее время нет основного способа проверки адекватности результатов тестирования. В связи с этим изучение способов оценки результатов тестирования является одним из главных вопросов в обучении и требует дальнейших исследований.

Адрес статьи: www.gramota.net/materials/1/2012/12-2/30.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2012. № 12 (67): в 2-х ч. Ч. II. С. 137-140. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2012/12-2/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

УДК 004/005

Технические науки

В данной статье рассматривается проблема анализа результатов в системах автоматического тестирования знаний. Автор дает обобщенную характеристику типам тестов, проводит сравнение алгоритмов тестирования, применяемых для автоматических тестов, указывает причины предпочтительности данного вида тестирования остальным. Выделяются и описываются характерные особенности основных алгоритмов, а также проблемы, которые могут возникнуть при использовании каждого из них. Автор обращает внимание, что в настоящее время нет основного способа проверки адекватности результатов тестирования. В связи с этим изучение способов оценки результатов тестирования является одним из главных вопросов в обучении и требует дальнейших исследований.

Ключевые слова и фразы: высшее образование; компьютерное тестирование; проверка; разработка тестовых заданий; алгоритмы проверки; инструменты тестирования; адекватность результатов.

Виктория Александровна Редюк*Кафедра вычислительной техники**Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики**victoriya.redjuk@comptel.com***СИСТЕМЫ АВТОМАТИЧЕСКОГО ТЕСТИРОВАНИЯ ЗНАНИЙ:
АЛГОРИТМЫ, ИНСТРУМЕНТЫ, АДЕКВАТНОСТЬ ПОЛУЧЕННЫХ ДАННЫХ[©]**

Вся система нашего образования, особенно среднего и высшего, построена на проверке знаний учащихся. Способов проверки существует большое количество. В последние несколько лет большую популярность набирает проверка знаний в форме тестирования. Особенно актуально это стало после введения обязательного ЕГЭ. Тестирование в своей первоначальной форме более просто для проверки знаний, чем остальные виды проверки. Но само проведение тестирования и проверка его результатов могут занять весьма продолжительное время. С этой точки зрения более предпочтительным оказывается компьютерное тестирование.

Компьютерное тестирование предполагает проведение проверки учащихся в форме теста, причем этот тест проходит с использованием каких-либо информационных технологий. В самом простом виде оно может представлять из себя какой-либо объект с заданием и вариантами ответов, среди которых есть один или несколько правильных. Учащийся имеет возможность выбора, этот выбор запоминается программой-тестом и после прохождения теста выдает результат, посчитанный по каким-то своим алгоритмам.

Одной из основных форм является форма закрытых ответов, при которой участнику на выбор предлагается несколько уже заранее сформированных ответов. Тестовые задания закрытого типа - самые распространенные в педагогике и психологии. Они сравнительно легко формируются, легче понимаются учащимися. Вместе с тем, закрытые тестовые задания могут иметь разнообразную внутреннюю структуру при сохранении общего дидактического смысла задания. Это - самый простой вид контроля, где программа проверяет правильность выбора ответа. Минусом такого тестирования может быть неграмотно сформированный ответ или несколько похоже сформированных ответов при возможности выборки только одного.

Второй наиболее распространенный вариант - это форма полуоткрытых ответов, она предполагает ввод ответа с клавиатуры в поле. Программа не предполагает выведенных заранее ответов, но в теории содержит максимально возможный, по мнению составляющих ее авторов, набор ответов. Считается, что в большинстве случаев в программе имеются необходимые варианты ответов, и она сможет грамотно сравнить ответ тестирующего с возможными вариантами в базе. Недосток такого варианта логичен - невозможно предугадать все варианты ответа тестирующего, и если он сделает грамматическую ошибку, то система посчитает ответ неверным. Имеются и другие варианты, и каждый из них имеет разные достоинства и недостатки.

Проведение тестирования, кроме своей простоты, имеет ряд существенных недостатков. Большая часть этих недостатков связана с разработкой тестовых заданий. Типы компьютерных тестовых заданий определяются способами однозначного распознавания ответных действий тестируемого.

1. Типы тестовых заданий по блоку «знания»:

- вопросы альтернативные (требуют ответа да-нет);
- вопросы с выбором (ответ из набора вариантов);
- вопросы информативные на знание фактов (где, когда, сколько);
- вопросы на знание фактов, имеющих формализованную структуру (в виде информационной модели или схемы знаний);
- вопросы по темам, где имеются однозначные общепринятые знаковые модели: математические формулы, законы, предикатные представления, таблицы;
- вопросы, ответы на которые можно контролировать по набору ключевых слов;

▪ вопросы, ответы на которые можно распознавать каким-либо методом однозначно.

2. Типы тестовых заданий по блоку «навыки» (распознавание деятельности: манипуляции с клавиатурой; по конечному результату):

- задания на стандартные алгоритмы (альтернативные да-нет, выбор из набора вариантов);
- выполнение действия.

3. Типы тестовых заданий по блоку «умения».

То же самое, что навыки, но используются нестандартные алгоритмы и задачи предметной области при контроле времени их решения:

- задания на нестандартные алгоритмы (альтернативные да-нет, выбор из набора вариантов);
- выполнение действия.

Разработка хороших тестов требует большой работы, чтобы все задания были ясно сформулированы и не допускали разночтений, кроме того требуется подобрать ответы, похожие на правду, чтобы учащийся не смог так просто исключить неправильные ответы.

В настоящее время существует много разных инструментов компьютерного тестирования. Они отличаются возможностями администрирования, добавления и проверки тестов, интерфейсом и т.п. Полный обзор таких систем выходит за рамки работы. Нас интересуют возможности системы компьютерного тестирования по проверке результатов тестирования.

Любая система тестирования имеет свои алгоритмы проверки теста. Есть четыре основных пути оценивания результатов, которые показывают более-менее адекватные результаты.

1. Сумма баллов. Оценка за тест - это сумма всех баллов за все вопросы. Это самый простой подход к оцениванию, и он не исправляет завышение или занижение балла при угадывании или случайной ошибке. Также на результате могут отражаться ошибки, допущенные непосредственно во время составления теста. Этот подход может быть использован только в случае, если уже был многократно выверен, а результаты суммирования баллов совпадают с результатами других видов тестирования.

2. Парное сравнение. Оценка, получаемая студентом, - это ранг суммы показателей сходства ответов рассматриваемого студента с другими тестируемыми. Результат сравнения студентов i и j может принимать три значения W_{ij} :

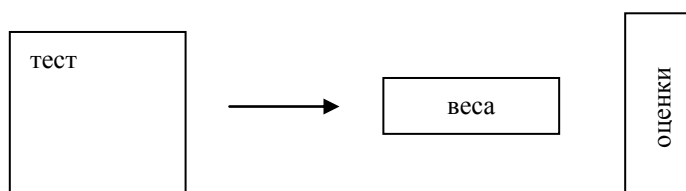
Результат сравнения		Возможные значения W_{ij}			
		вариант 1	вариант 2	вариант 3	метод Саати
i лучше j	a	2	1	1	2-9
i равно j	b	1	0.5	0	1
i хуже j	c	0	0	-1	1/2-1/9

Процедура выполняется в 2 шага. На основе результатов тестирования строится таблица сравнений каждой пары студентов списка, а потом оценки сходства суммируются для каждого студента (Сумма) и сортируются по убыванию ранга.

	Первый	Второй	Третий	Сумма	Ранг суммы
Первый	1	2	2	5	1
Второй	0	1	2	3	2
Третий	0	0	1	1	3
....					

Помимо суммирования и ранжирования иногда выясняют средний арифметический показатель или среднее геометрическое и рассчитывают долю оценки в сумме всех оценок. Подобная технология дает возможность представления индивидуальной оценки знаний, взвешенной относительно всех оценок знаний всех остальных студентов. Однако такой способ не дает возможности проанализировать характеристику или значимость отдельных вопросов теста.

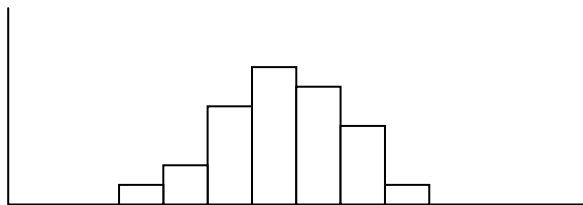
3. Факторный анализ или анализ компонентов базовой таблицы оценок. Этот метод анализирует корреляции ответов и позволяет получить численные выражения «весов» каждого из вопросов в виде вектора факторных нагрузок. Также метод позволяет получить рейтинговые оценки знаний студентов в форме вектора значимый компонент (факторов). Если все вопросы похожи по информативности, то получают единственную значимую главную компоненту или вектор, который совпадет с простой суммой баллов. Если вопросы неодинаковы и могут быть сгруппированы по степени информативности, то компонентный анализ может выдать несколько различных оценок знаний, каждая из которых будет основана на какой-либо группе вопросов.



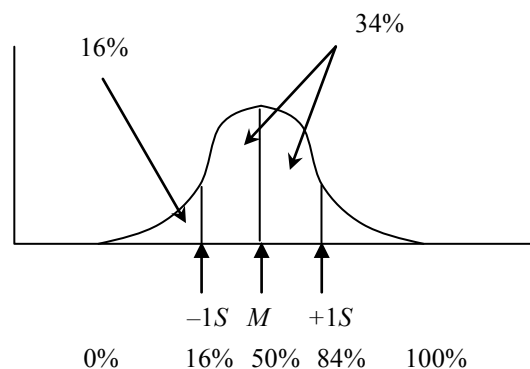
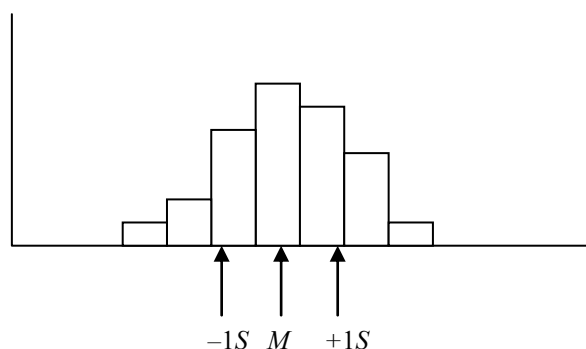
4. Многомерное шкалирование. Результат аналогичен результату корреляционного анализа. Различие состоит в том, что в компонентном анализе используется коэффициент корреляции, а в многомерном шкалировании в основном - обычные показатели разнообразия (мера Шеннона, Жаккара).

Интегрированные оценки, полученные при помощи одного из названных методов, используются для создания ранговой шкалы знаний студентов и, в итоге, для создания правила, кому какие оценки ставить.

Строится ось координат OX со значениями интегральных оценок знаний.



Затем строится распределение студентов по интервалам оценок.



Находится среднее арифметическое M и стандартное отклонение S , выраженное в баллах.

Далее на диаграмме отмечаются три значения - $M-S$, M , $M+S$, и распределение делится на пять областей - от 0 до 16, от 16 до 50, от 50 до 84 и от 84 до 100%, где каждый из отрезков соответствует оценке по 5-балльной шкале. В соответствии с положением оценки студента, выставляется оценка в привычной для студента форме.

Для верификации теста используются мнения сторонних экспертов, проводят свой собственный опрос для той же группы и затем сравнивают оба варианта результатов. Чем больше привлекается экспертов, тем объективнее считается оценка и проверка.

Описанные процедуры используются в основном на этапе создания и доработки теста, тогда как на этапе выставления ограничиваются подсчетом и ранжированием сумм баллов, а затем переводят в привычные оценки по пятибалльной шкале.

При объективном анализе полученных результатов в поле зрения попадает не просто совокупность ответов на вопросы, но и относительная сложность вопросов. Несмотря на постоянные попытки создавать более менее равные вопросы, все равно какие-то будут сложнее, а какие-то легче. Кроме того, не исключена возможность списывания теста учащимся. Это искажает результаты теста и приводит к необходимости учета вероятностной составляющей при их анализе или может привести к ситуации, когда учащийся может ошибочно попасть под подозрение на предмет взлома или обмана системы.

Многие способы анализа результатов учитывают подобные ситуации и назначают вопросам определенные веса, которые в теории могут предотвратить частичные ошибки сильных студентов или угадывание ответа слабыми. К сожалению, в настоящее время нет единого и надежного способа проверки адекватности результатов тестирования. Поэтому одним из главных вопросов является изучение способов оценки адекватности результатов тестирования.

В данный момент времени в качестве алгоритмов оценки используются различные статистические и статистически-временные алгоритмы, когда оценка результата оценивается с помощью различных средних результатов учащихся или его конкретных характеристик. Такие алгоритмы достаточно просты для программирования, но могут потребовать знания некоторых характеристик учащихся или задач, которые могут быть неизвестны на начало теста. Кроме того, возможно построение простой нейронной сети, и на основе некоторых тестов последующее обучение данной сети выдавать адекватную оценку. Здесь возникают проблемы с обучением сети, какой алгоритм обучения использовать, и как выбирать тесты для обучения сети.

Список литературы

1. Аванесов В. С. Тесты: теория и методика их разработки. Ст. 1-10 // Управление школой. 1999. № 36. С. 32-33.
2. Бабанский Ю. К. Педагогика: учебное пособие для студентов пед. ин-тов. М.: Просвещение, 1988.

3. **Оплачко Е. С.** Искусственные нейронные сети. Проблемы и перспективы развития // Альманах современной науки и образования. Тамбов: Грамота, 2010. № 9. С. 47-49.
4. **Семушина Л. Г., Ярошенко Н. Г.** Содержание и методы обучения в средних специальных учебных заведениях. М.: Высшая школа, 1990.
5. **Эрганова Н. Е.** Методика профессионального обучения: учеб. пособие. Екатеринбург: Рос. гос. проф.-пед. ун-т, 2003. 150 с.
6. http://ifets.ieee.org/russian/depository/v13_i3/html/6r.htm
7. <http://u4im-u4im.ru/79-kompyuternoe-testirovanie-znanij.html>

УДК 338.24:005.336.3:378

Экономические науки

В статье представлены результаты исследования особенностей формирования процессно-ориентированной системы управления качеством услуг высшего образования. Графически изображена система процессно-ориентированного управления качеством услуг вуза. Определены ключевые характеристики процессов управления в сфере высшего образования. Предложены методические основы разработки процессно-ориентированных систем управления качеством в сфере высшего образования.

Ключевые слова и фразы: управление качеством; высшее образование; процесс; система.

Алексей Владимирович Родионов, к.э.н.

Кафедра экономики предприятия и управления трудовыми ресурсами

Луганский национальный аграрный университет, Украина

a.v.rodionov@list.ru

ФОРМИРОВАНИЕ ПРОЦЕССНО-ОРИЕНТИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ КАЧЕСТВОМ УСЛУГ ВЫСШЕГО ОБРАЗОВАНИЯ[©]

Управление качеством услуг вуза на основе процессного подхода - одно из ключевых требований стандартов ISO, которые являются базовой и общеметодологической основой применения большинства специфических образовательных стандартов управления качеством. Стандарты ISO являются документально и нормативно закрепленным выражением философии процессно-ориентированного управления, при этом в пределах отмеченной концепции качества это управление является тотальным (*TQM, Total Quality Management*, англ. *Тотальное управление качеством*), то есть объектом управления является вся система предоставления услуг со стадиями, предшествующими их предоставлению, и стадиями, которые имеют место по окончании процесса предоставления образовательной услуги. При этом весь процесс производства и предоставления услуги рассматривается как систематизированная совокупность связанных процессов более конкретного уровня, которые, в свою очередь, могут также иметь атомарную структуру.

В целом, понятие процесса является фундаментальным как для современного менеджмента, так и для менеджмента качества в частности. Под процессом в самом широком смысле понимается некоторая последовательность взаимосвязанных или взаимодействующих действий (работ, операций), целью которой является превращение «входов» процесса в его «выходы» для достижения определенного результата, как правило - создания некоторой продукции или предоставления какой либо услуги для потребителей [1, с. 32; 4, с. 58].

В пределах процесса среди объектов управления чаще всего выделяют определенные ресурсы, которые разделяют на ресурсы долгосрочного и краткосрочного использования. При этом под ресурсом обобщенно понимается широкий круг факторов производства и объектов материально-технического обеспечения, которые используются при предоставлении образовательных услуг: профессорско-преподавательский состав, персонал, здания и сооружения, технологии и методы обучения, оборудование, техника, объекты интеллектуальной собственности и т.д. Каждый из процессов рассматривается как отдельная система с определенным уровнем автономии, как следствие, в анализе нуждается внешняя среда процесса, которая оказывает влияние на него. Но наиболее важным вопросом является то, что каждый из процессов должен быть управляемым. Иначе говоря, каждый из процессов является зоной конкретных управленческих полномочий и ответственности. Используя современную философию управления качеством, должностное лицо, ответственное за реализацию отдельного процесса системы управления качеством, является владельцем процесса или его руководителем.

При рассмотрении деятельности организаций традиционно говорят о рабочих процессах или бизнес-процессах. Последнее понятие в настоящее время используется применительно к любым организациям, не обязательно промышленным, включая вузы. Одним из основных необходимых свойств любого рабочего процесса является его так называемая «способность» (англ. *Capability*), под которой понимается способность процесса выполнять нужные функции и производить на выходе результат или услуги с установленными и ожидаемыми свойствами. Для обеспечения этого свойства в схему процесса обычно включается