

ОБЪЕКТНО–ИЕРАРХИЧЕСКАЯ МОДЕЛЬ ДОСТОВЕРНОСТИ ТЕСТОВОГО КОНТРОЛЯ

Светлана Кудинова, Александр Митрошин, Андрей Митрошин

Криворожский филиал Европейского университета
Украина, 50068, г Кривой Рог, ул. Галенка 2
Тел.: (0564) 404-03-83, 92-11-54, E-mail: info@kdeu.dp.ua

Abstract

One of the possible approach is offered In work to practical entailment of the theoretical achievements in practical person of the development test. They Are Considered practical aspects of the reception of the qualitative features test, and is formulated object-hierarchical model of the test checking with procedure of the estimation.

Анализ источников и постановка задачи исследований

Рассмотрим суть проблемы тестирования и оценки качества тестов. При тестировании респондентов, с целью выявления уровня их знаний полученное число баллов зависит от сложности тестовых задач и уровня знаний респондента одновременно. В общем случае можно говорить о двухмерной функции зависимости вероятности правильного ответа от сложности вопроса и уровня знаний респондента. Суммарный полученный балл содержит, как некоторую закономерную составляющую, так и случайную. Выбор функциональной зависимости вероятности правильного ответа от характеристик вопроса и респондента произволен. Поэтому существует значительное число разнообразных моделей[1], которые зависят от ряда параметров. Для описания вероятности правильного ответа i -го респондента на j -й вопрос используем функцию Раша [1]:

$$P_{i,j} = \frac{e^{Dv_j(a_i - Q_j)}}{1 + e^{Dv_j(a_i - Q_j)}}, \text{ при } Q_j \rightarrow -\infty$$

где: a_i - уровень знаний респондента; Q_j - уровень сложности вопроса; v_j - коэффициент дифференциации знаний вопросом; $D=1.7$ - числовой параметр, который нормирует шкалы вероятностей для различных математических моделей. Коэффициенты a_i и Q_j являются латентными признаками, которые непосредственно не наблюдаются. Процедура определения латентных признаков требует объемных тестовых экспериментов с оцениванием уровня знаний и уровня сложности вопросов. Эти показатели можно измерять в соответствующих безусловных вероятностях, или в логитах. Если a_i - вероятность, то логит получим согласно выражения $\ln\left(\frac{a_i}{1-a_i}\right)$. Обратное преобразование следующее $e^{a_i}/(1+e^{a_i})$.

При статистической обработке результатов тестирования лучше пользоваться вероятностями оценок, поскольку нелинейное преобразование в логиты приведет к ненормальности распределения случайной составляющей. Это усложнит построение доверительных интервалов для оценок. Нормальность случайной составляющей в результатах тестирования следует из центральной предельной теоремы. Рассмотренная модель служит основой, с помощью которой исследуем алгоритмы оценки характеристик теста. Для моделирования процесса тестирования и исследования алгоритмов обработки результатов тестирования разработан программный модуль «tests» в среде MATLAB [1].

Сложность вопросов и уровень знаний респондентов задаются случайно и считаются распределенными по нормальному закону. Верный ответ формируется как случайное событие в одном опыте, имеющее биномиальный закон распределения. Обработка результатов состоит в определении средних баллов респондентов, которые являются оценками уровня знаний. Средний процент верных ответов по вопросам служит оценкой сложности вопросов. Средний процент верных ответов на вопрос респондентов с одинаковыми оценками уровня знаний служит оценками вероятности верного ответа на отдельные вопросы теста респондентов с известными оценками уровня знаний. Последние точки образуют профили тестовых вопросов. Полученный эмпирический профиль неизвестен и содержит значительную шумовую составляющую.

Построим профили тестовых вопросов по отдельным эмпирическим точкам, т.е. решим задачу сглаживания эмпирических данных некоторой функцией. Выбор функции значительной мерой произволен и основан на интуиции и опыте оператора, а это затрудняет автоматизацию этой процедуры. При применении функции Раша возникают трудности оценки параметров, которые входят в уравнение нелинейно. Решение задачи. Используем функции[2], которые имеют хорошую адаптивную способность в

сочетании с простотой расчетов и линейной зависимостью от параметров. В практике обработки данных наиболее известны интерполяционные кубические сплайны с двумя непрерывными производными [3,4].

```
% моделирования процедуры тестирования дихотомическим тестом
% и обработки результатов с выявлением латентных признаков
N=300; % количество респондентов
M=20; % количество вопросов
pr=normrnd(0,1,N,1); % уровень знаний респондентов (в логитах)
pt=normrnd(0,1,M,1); % уровень сложности вопросов (в логитах)
for r=1:N
    % моделирования тестирования
    for t=1:M
        % вероятность верного ответа (по Рашу)
        p=exp(1.7*(pr(r)-pt(t)))/(1+exp(1.7*(pr(r)-pt(t))));
        x(r,t)=binornd(1,p,1,1); % моделирования ответа
    end
end
% обработка результатов тестирования
for r=1:N
    xpr(r)=sum(x(r,:)); % сумма баллов респондента
ppr(r)=xpr(r)/M; % средний балл респондента (вероятность правильных ответов)
end
for t=1:M
    xpt(t)=sum(x(:,t)); % сумма баллов вопроса
    ppt(t)=xpt(t)/N; % средний балл вопроса (вероятность правильных ответов )
end
qp=zeros(M+1,M); % число баллов(оценка знаний, номер вопроса)
np=zeros(M+1,M); % число респондентов(оценка знаний, номер вопроса)
for r=1:N
    for t=1:M
        qp(xpr(r)+1,t)=qp(xpr(r)+1,t)+x(r,t);
        np(xpr(r)+1,t)=np(xpr(r)+1,t)+1;
    end
end
Q=zeros(M+1,M); % оценка знаний(оценка знаний, номер вопроса)
for i=1:M+1
    if np(i)>0
        qp(i,:)=qp(i,:)/np(i,:); % оценка вероятности верного ответа
        % на вопрос
    end
    Q(i,:)=i;
end
prp=exp(pr)/(1+exp(pr)); % уровень знаний респондентов (в вероятностях)
ptr=exp(pt)/(1+exp(pt)); % уровень сложности вопросов (в вероятностях)
```

Рис.1. Листинг программы «tests» в среде MATLAB

Для их расчета необходимо решить систему интерполяционных уравнений. В данной работе использованы кубические эрмитовые сплайны, которые имеют непрерывность лишь первой производной и локальны. При интерполяции с их помощью нет необходимости решать систему уравнений. Именно такая модель была применена в работе для описания профиля тестовых вопросов. Для оценивания вектора A - параметров сплайновой модели, был применен метод наименьших квадратов (МНК). Такой выбор обусловлен нормальностью случайной составляющей и простотой метода. Оцениваемыми параметрами являются значения сплайна в точках стыковки фрагментов. Особенность сплайнов в существенной зависимости их формы от количества и схемы размещения точек стыковки – узлов: (tu, A) . Поэтому алгоритм построения сплайна в точках стыковки должен учитывать это обстоятельство. Есть два варианта решения этого вопроса. Рассмотрим первый вариант решения вопроса - последовательное построение сплайна с контролем отсутствия неслучайной составляющей в остатках приближения. Понятно, что проверка случайности имеет характер проверки статистической гипотезы, для чего используем непараметрический критерий инверсий. Очередной фрагмент сплайна продолжается пока не будет отброшена гипотеза о случайности остатков. Достоинством предложенного алгоритма есть автоматическое определение числа узлов. Вероятность ошибки первого рода следует выбирать таким образом, чтобы вероятность случайно-

сти остатков была максимальной. Последовательное построение сплайна по МНК реализует модуль «sersplm»:

```
alfa=0.5; % уровень ошибки первого рода
[tu,A,Am,Ap,PInv]=sersplm(Q(:,nt)/(M+1),qp(:,nt),0,alfa); % последовательное построение
sp=splherm(r,tu,A); % получение графика оцененного профиля
p=rashp(r,ptp(nt)); % получение истинного профиля
plot(r,p,r,sp,tu,Am,'o',tu,Ap,'o',Q(:,nt)/(M+1),qp(:,nt),'+');title([' quest ',num2str(nt)]);
```

Для получения «МНК оценок» при заданной схеме размещения узлов воспользуемся модулем «lsqsplh», который вместе с оценками рассчитывает и доверительные интервалы для узлов и точек наблюдений:

```
[A,di,s,ds]=lsqsplh(tu,Q(:,10)/(M+1),qp(:,10),0.95);
```

Второй вариант решения вопроса состоит в применении метода покоординатного спуска с ограничениями. При заданной начальной сетке узлов на каждом этапе оптимизации находится оптимальное положение узла относительно соседних. Оптимизация осуществляется по критерию минимума среднеквадратичного отклонения. Начальное положение узлов можно найти путем последовательного построения, а покоординатной оптимизацией их уточнить. Вызов модуля «optspl» покоординатной оптимизации:

```
[wu,wi,op]=optspl(tu,Q(:,4)/(M+1),qp(:,4),4);
```

На рис. 2 показаны теоретические профили трех тестовых вопросов со сложностями 0.238, 0.57, 0.90, соответственно и эмпирические результаты моделирования процесса тестирования. Результаты компьютерной обработки модельных результатов показаны на рис. 3, 4, 5.

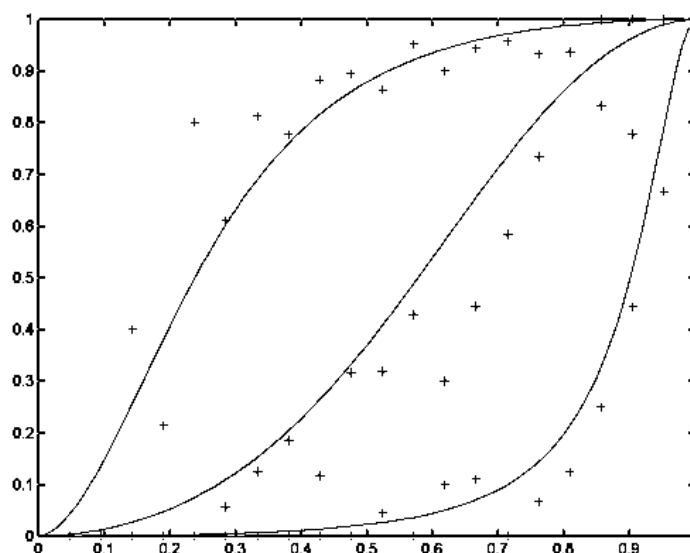


Рис.2. Теоретические профили трех тестовых вопросов со сложностями 0.238, 0.57, 0.90

Сплайновая модель точно передает характерные особенности профилей тестов с разными уровнями сложности и при этом обеспечивается практически достаточный уровень достоверности. Другим важным аспектом идентификации модели является компактное представление модели профиля тестовых вопросов. Вся информация о модели оценивания содержится в значениях узловых точек сплайна. Достоверность полученной математической модели описывается шириной доверительных интервалов в точках стыковки. Представленные алгоритмы оценивания характеристик вопросов общие и не учитывают некоторых особенностей обрабатываемых данных. Их дальнейшее усовершенствование состоит в применении взвешенного МНК, (данные являются оценками частоты с разным объемом выборки) и учете ограничений решения (в пределах от 0 до 1).

Оценка профилей тестовых вопросов требует значительного числа респондентов с разным уровнем знаний, поэтому целесообразно использовать итерационную процедуру оценки характеристик теста. Первый этап этой процедуры состоит в проектировании и разработке теста с проведением начального тестирования. Методика проведения такого анализа известна. На этом этапе анализа выявляют наиболее очевидные недостатки и устанавливаются общие характеристики на некоторой контрольной группе респондентов. Для ускорения процесса уточнения характеристик теста и снижения стоимости этой процедуры были привлечены к этому процессу широкий круг тестируемых. Процедура аналогичная бета тестированию программного обеспечения. Тесты передавались бета тестируемым на условиях, которые предусматривают их заинтересованность и ответственность. Оперативную связь осуществлялась через электронную почту и интерактивные страницы. К разработчикам теста поступала информация о характеристиках тестовых групп и результаты тестирования.

После достижения необходимого уровня достоверности оценок профилей тестовых вопросов выполнялась оптимизация теста (изъятие малоинформативных вопросов и вопросов со статистически близкими профилями). Сформировался паспорт теста, который включает профили вопросов.

Наличие этих характеристик дает возможность более объективно оценивать уровень знаний респондентов. С этого момента тест считается завершенным информационным продуктом с нормированными характеристиками.

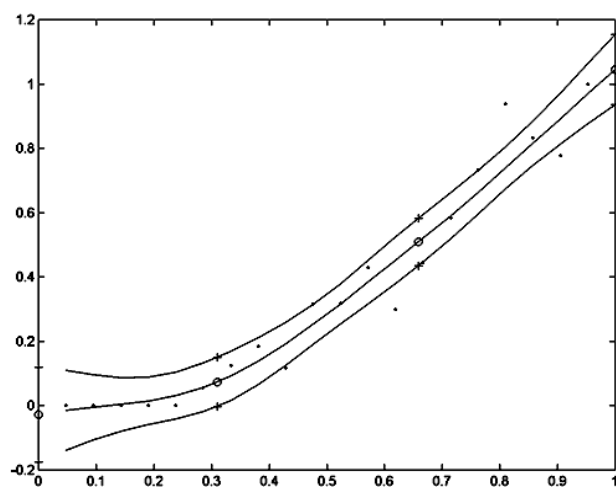


Рис. 3. Оценка профиля вопроса сложности 0.57.

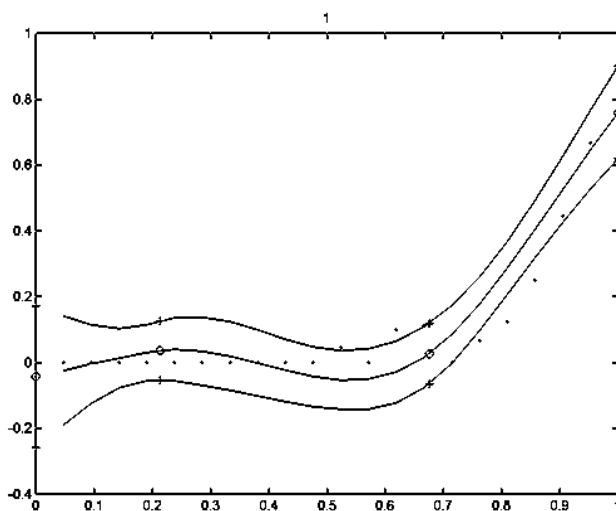


Рис. 4. Оценка профиля вопроса сложности 0.24.

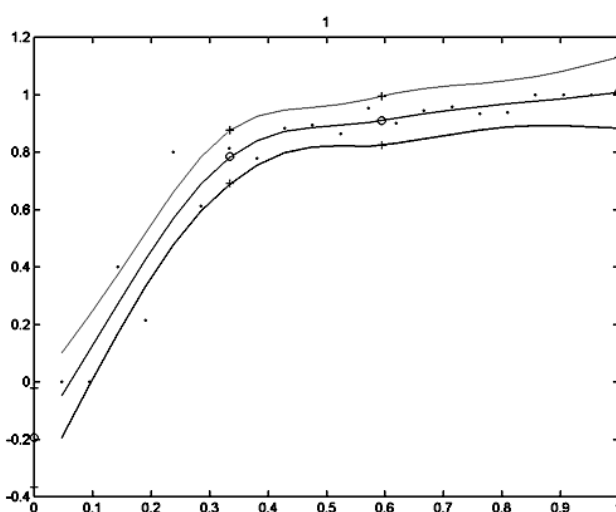


Рис. 5. Оценка профиля вопроса сложности 0.90.

В настоящее время возникает проблема установления определенных стандартов на такие измерительные инструменты [5]. С технической точки зрения тестовый инструмент может основываться на иерархической структуре объектов, которая должна иметь следующую структуру: тест: область знаний

(например, за УДК), предназначение теста, категория респондентов, e-mail адреса авторов, список авторов, аннотация авторов, дата создания, дата просмотра, условия лицензии, вопросы; вопрос: категория, группа, текст вопроса, варианты ответов, верный вариант ответа, оценка профиля вопроса; группа респондентов: дата тестирования, реквизиты ответственного за тестирование, аннотация группы, априорная оценка уровня знаний группы, первичное или повторное тестирования; респондент: идентификатор респондента, априорный уровень знаний, профессиональные реквизиты респондента, массив ответов на тестовые вопросы, суммарный балл, оценка уровня знаний респондента, оценка профиля респондента.

Выводы. На основании проведенных исследований предложен метод формирования, испытания и применения тестов способный гарантировать необходимый уровень достоверности систем оценки знаний, превратить тест в измерительный инструмент с нормированными характеристиками относительно погрешностей.

Література:

- [1] Доровской В.А. Дихотомическое тестирование и оценивание эффективности обучения диспетчеров полиэргатических систем. – Днепропетровск: Наука і освіта, 2001. –412 с.
- [2] Приближение функций для технических приложений / Б. А. Попов, Г. С. Теслер. - Киев : Наук. мысль, 1980. - 350 с.
- [3] Корнейчук Н.П. Сплаины в теории приближения.- Г.: Наука, 1984. - 352с
- [4] Комплекс нормативних документів для розробки складових систем стандартів вищої освіти. Додаток 1 до наказу Міносвіти N 285 від 31 липня 1998 р.