

Методология

ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ ТЕОРИИ ИНФОРМАЦИИ И ТЕОРИИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Юрий Каргин,

Ноябрьский колледж профессиональных
и информационных технологий
kargin04@yandex.ru

В статье рассмотрены некоторые вопросы применения теории информации к педагогическим измерениям. Такой подход позволил привести информационную интерпретацию основных показателей моделей педагогических измерений, выделить совокупность моделей Раша и построить алгоритм их конструирования, а также предложить методы оценки эффективности тестовых испытаний.

Ключевые слова: педагогические измерения, модель педагогических измерений, единицы педагогических измерений, IRT, формула Шеннона

Введение

Развитие любой науки связано с введением собственных понятий и величин, на которых она базируется. Не является исключением и математическая теория измерений (IRT). Во всём множестве смежных и собственных понятий IRT¹ некоторые из них имеют исключительное значение. К таким относятся количественные показатели мер подготовленности испытуемого b (person ability, обозначения по работе²) и трудности тестового задания d (item

1

Аванесов. В.С.
Item Response Theory:
основные понятия и по-
ложения. ПИ №2. 2007.
С. 3–28;
Истоки и основные по-
нятия математической
теории измерений (Item
Response Theory). Ста-
тья вторая. ПИ №3.
2007. С. 3–36.

2

*Wright B.D.,
Stone M.H.*
Measurement Essentials.
WIDE RANGE, INC.
Wilmington, Delaware,
1999. 221 p.

difficulty), значения которых определяются по латентной шкале логитов. Причём введённая в IRT единица измерения логит является внутренней единицей измерений и никак непосредственно не связана с универсальными или широко распространёнными и в других отраслях знаний единицами измерений.

Содержательная интерпретация значений показателей b и d , точнее расстояний между ними $g = b - d$ сводятся к описанию зависимости от этой разности доли правильных ответов P , информационной функции $I = P \cdot (1 - P)$ и некоторых менее значимых показателей. Фактически в различных изложениях IRT именно эти рассуждения, иногда подкреплённые графиками, дают необходимый содержательный смысл о результатах измерений.

В данной работе предлагается расширить содержательный смысл основных показателей педагогических измерений и связать основные единицы педагогических измерений с информационными мерами. Основания для этого можно найти в широко известной формуле Шеннона.

Выявленная связь формулы Шеннона с математической моделью педагогических измерений Раша позволила также предложить алгоритм констру-

ирования математической модели педагогических измерений по форме тестового задания. В качестве примера сконструированы как некоторые известные, так и частные модели педагогических измерений.

И, наконец, информационные меры можно использовать при оценивании эффективности теста. Эти вопросы нашли своё отражение в завершающих подразделах данной работы.

Информационная энтропия Шеннона

Существуют различные понятия энтропии. Мы используем это понятие в формулировке Шеннона. Клод Шеннон предложил количественно рассчитывать полученную от элементарного сообщения информацию как меру снятой неопределённости (энтропии) этого сообщения. Неопределённость одного состояния x_i сообщения x называют частной энтропией, её значение рассчитывается по формуле

$$h(x_i) = -\log_2 P_i,$$

где P_i — вероятность появления состояния из набора дискретных состояний элементарного сообщения, $\sum P_i = 1$. Энтропия всего элементарного сообщения определяется как среднее частных энтропий по всем состояниям: $H(x) = -\sum P_i \cdot h(x_i)$.

Единицей измерения энтропии является бит³ (binary digit – двоичное число).

В данной работе осуществляется опора преимущественно на формулы Шеннона с натуральным логарифмом – для частной энтропии $h(x_i) = -\ln P_i$ и для средней энтропии $H(x) = -\sum P_i \cdot \ln P_i$.

При расчёте энтропий по этой формуле результат имеет единицу измерения *нат* (nat).

Если случайное событие x имеет только два состояния (1 – да; 0 – нет) с вероятностями P и $1 - P$, то формулы Шеннона приобретают вид:

$$\begin{aligned} h(x=1) &\equiv h_1 = -\ln P, \\ h(x=0) &\equiv h_0 = -\ln(1-P), \\ H(x) &= -P \cdot \ln P - (1-P) \ln(1-P). \end{aligned}$$

Кратко остановимся на свойствах этих функций.

Графики функции $h(x)$ и $H(x)$ приведены на рис. 1а и 1б.

Графики частных энтропий монотонные убывающая для h_1 и возрастающая для h_0 функции с положительной областью

значений от нуля до бесконечности и вертикальными асимптотами соответственно $P = 0$ и $P = 1$. При $P = 0,5$ значения частных энтропий совпадают и равны $h_1(0,5) = h_0(0,5) = \ln 2$.

Энтропия H ограничена и достигает максимума, равного $\ln 2$ нат (или 1 бит) при $P = 0,5$. При $P = 0$ или $P = 1$ неопределённость H равна нулю. Среднее значение функции $H(x)$ обозначено на рисунке горизонтальной пунктирной линией и равно $H_{\text{ср}} = 0,5$ нат, что примерно соответствует неопределённости сообщения с вероятностями событий $P = 0,2$ или $P = 0,8$. Продолжим исследование этой функции. Вычисляя производную, находим

$$\frac{dH}{dP} = \ln \frac{1-P}{P} = h_1 - h_0,$$

т.е. производная энтропии равна разности частных энтропий благоприятного и неблагоприятного состояний.

Последнее выражение, наряду с определениями функ-

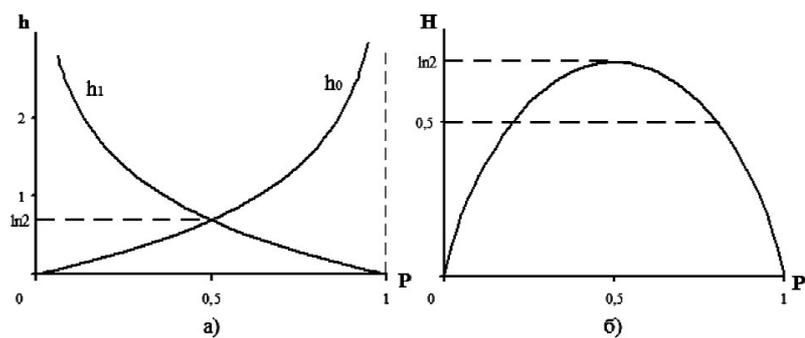


Рис. 1. Графики функций энтропий $h(x)$ и $H(x)$

Методология

3

Двоичная единица измерения информации «бит» – не единственная мера. Эта мера определяется основанием логарифма – 2. Если в формуле Шеннона изменить двоичный логарифм на натуральный, то такую единицу измерения называют «нат». Связь между этими единицами измерения следующая: 1 бит = $\ln 2 \cdot 1$ нат. Безусловно, количество информации от выбора основания логарифма не зависит, выбор основания и, как следствие, выбор единицы измерения – лишь вопрос масштаба измерительной шкалы. Единственное требование – в одном тесте для результатов измерений должна использоваться единая шкала. При формальных математических преобразованиях можно считать величину количества информации безразмерной (как, впрочем, и логит).

ПЕД
измерения

ций энтропии, отражает связь производной усреднённой энтропии через разность частных энтропий симметричных состояний.

Для дальнейших рассуждений важно отметить и то, что производная энтропии

$$\frac{dH}{dP} = -\ln \frac{P}{1-P}$$
 с точностью до

знака совпадает с выражением для исходной оценки логита уровня подготовленности испытуемого

$$b = \ln \frac{P}{1-P}$$
 в модели

Раша. Такое совпадение даёт возможность не просто выразить показатели педагогических измерений модели Раша через информационные показатели энтропии Шеннона, но и попытаться выявить содержательные связи между соответствующими теоретическими положениями.

Действительно, даже рассуждая формально, модель Раша можно вывести из формулы Шеннона. Если в выражении для производной функции Шеннона

$$\frac{dH}{dP} = -\ln \frac{P}{1-P}$$
 вести обозначение $g = -\frac{dH}{dP}$ (или $g = h_0 - h_1$)

и выразить вероятность P , то получим формулу Раша $P = \frac{\exp g}{1 + \exp g}$ для вероятности правильного ответа испытуе-

мым с подготовленностью b при выполнении задания трудности d , где $g = b - d$. График этой функции хорошо известен в ИРТ (рис. 2). Иными словами, Функция Раша (логистическая кривая) представляет собой график производной информационной энтропии Шеннона, взятой с противоположным знаком.

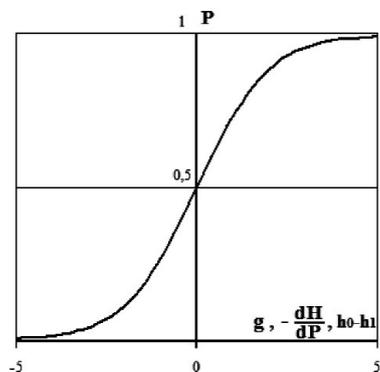


Рис. 2. Логистическая кривая с различными осями абсцисс

Таким образом, приведённые понятия и свойства функций энтропии Шеннона позволяют установить функциональную связь информационных показателей h и H , измеряемых в единицах нат, с измеряемыми в логитах значениями параметра $g = b - d$ модели Раша.

Пример

Рассмотрим пример, уточняющий понятия неопределённости. Учитель, знающий своего

ученика, при контроле знаний стремится дать ему адекватное задание или, как говорится, задание по уровню его развития. При выполнении такого задания и поиске правильного ответа на поставленный вопрос ученик должен испытывать существенные затруднения. Как он ответит на это задание, предсказать достаточно трудно; неопределённость состояний правильного или неправильного ответа ученика значима, неопределённость результата такого испытания близка к максимальной. Именно ответ на такое задание даст учителю наибольшее количество полезной информации об уровне его подготовленности.

Несложно привести примеры и неудачных заданий. Бессмысленно давать очень лёгкие задания. Неопределённость правильного ответа на него близка к нулю. Скорее всего, испытуемый легко выполнит задание, и этот предсказуемый результат не даст учителю много полезной информации об уровне подготовленности ученика. Так же совершенно бессмысленно задавать испытуемым очень трудное задание. Ждать от них правильные ответы не приходится, результат такого контроля знаний обычно предопределён. Ответы на такие задания дополнительной информации о подготовленности тестируемых дают мало.

Формулы Шеннона

Эти качественные рассуждения пока никак не подкреплены количественными оценками. Однако возможно и количественное описание подобных ситуаций. Такую возможность дают формулы Шеннона. Действительно, для эффективного задания вероятность правильного ответа близка к значению $P = 0,5$, неопределённость состояний правильного h_1 или неправильного h_0 ответа ученика оцениваются значениями $\ln 2$, неопределённость такого результата H близка к максимальной. Ответ ученика на задание снимает эту неопределённость и даёт информацию в количестве снятой неопределённости.

Для задания с хорошо предсказуемым результатом вероятность правильного ответа близка к единице (или нулю), частная энтропия такого состояния, как и средняя энтропия, приближаются к нулю, результат такого испытания информации практически не несёт. Эффективность такого задания близка к нулю уже не в качественном смысле, а в строго количественном.

Этим примером мы попытались раскрыть прикладной смысл понятий энтропии, значение математических формул Шеннона при расчёте неопределённостей и трактовку результатов их применения.

Информационная трактовка модели Раша

Попытаемся понять логику педагогических измерений, заложенную в формуле Раша, с позиции теории информации. Для удобства изложения этого вопроса изобразим графики функций $P(h)$, обратных частным энтропиям $h_0(P)$ и $h_1(P)$ (рис. 3а):

Первым рассмотрим случай с вероятностью $P = 0,5$ правильного выполнения испытуемым задания (точка А на рис. 3). Неопределённость состояний «правильный ответ» и «неправильный ответ» совпадают, и равны $h_1 = h_0 = -\ln 0,5 = \ln 2 \approx 0,693$, разность частных энтропий равна нулю $h_0 - h_1 = 0$. Это состояние соответствует равенству уровней подготовленности испытуемого b и трудности тестового задания d , т.е. нуле-

вой разности этих уровней $g = b - d = 0$.

Рассмотрим другой случай (обозначение на рисунках точкой В). Испытуемый достаточно хорошо подготовлен к выполнению тестового задания и правильно отвечает на его вопрос с вероятностью 0,8. Неопределённость состояния «испытуемый правильно выполнит задание» снижается, и частная энтропия этого состояния примерно равна $h_1 = -\ln 0,8 \approx 0,223$. При этом энтропия противоположного события «испытуемый неправильно выполнит задание» возрастает и примерно равна $h_0 = -\ln 0,2 \approx 1,609$. Разность частных энтропий этих состояний равна $h_0 - h_1 = 1,386$, т.е. неопределённость состояния «неправильный ответ» на 1,386 нат выше неопределённости состояния «правильный ответ». Эта разница выделена на рис. 3а фигурной скобкой.

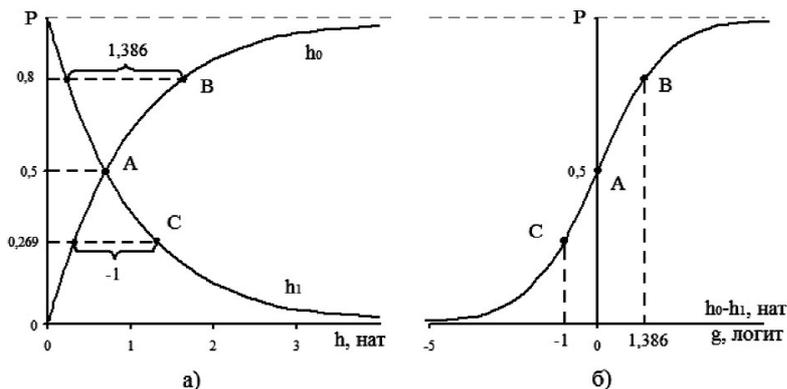


Рис. 3. Информационная трактовка модели Раша: частная энтропия h

Если учесть ранее установленную связь $g = h_0 - h_1 = \ln \frac{P}{1-P}$,

то такой результат в традиционной трактовке формулируется так: уровень подготовленности испытуемого на 1,386 логит выше уровня трудности тестового задания.

В третьем случае испытуемый недостаточно хорошо подготовлен к выполнению тестового задания и правильно отвечает на него с вероятностью 0,269 (точка С). Для такого состояния имеем разность энтропий $h_0 - h_1 = -\ln 0,731 + \ln 0,269 = 0,313 - 1,313 = -1$ нат, или $g = -1$ логит и вывод — уровень подготовленности испытуемого на один логит ниже уровня трудности тестового задания.

Если на рис. 3а от оси абсцисс частных энтропий h перейти к оси абсцисс разности энтропий $h_0 - h_1$, то предыдущие примеры расчётов будут представлены точками на традиционной т.н. логистической кривой (рис. 3б).

Отсюда первое заключение: традиционный показатель измерительной системы Раша $g = b - d$, измеряемый в частных научных единицах измерения логит, есть не что иное, как разность частных энтропий неправильного и правильного ответов на тестовое задание $h_{01} = h_0 - h_1$, измеряемых в об-

щенаучных единицах количества информации нат.

Дополнительное информационное содержание основных показателей педагогических измерений можно найти и в свойствах средней энтропии $H(x)$. Выше мы установили фор-

мальное соотношение $g = -\frac{dH}{dP}$.

Попытаемся понять, какой содержательный смысл вносит производная энтропии в толковании показателей педагогических измерений.

Аналитически, производная $\frac{dH}{dP}$ выражает скорость измене-

ния функции. Геометрически производная представляет собой тангенс угла наклона касательной к графику функции. В естествознании подобные выражения иногда обозначают термином «плотность» или «интенсивность». Можно опереться на любой из этих подходов, но в данном случае удобно начать с процентной интерпретации, иногда используемой в экономических приложениях.

Выражая приращение энтропии dH через приращение dP , получим $dH = -g \cdot dP$. Последнюю формулу можно приближённо трактовать так: рост вероятности правильного ответа P на 1% (от единичной вероятности достоверного события) приводит к уменьшению

ПЕД
измерения

энтропии системы на величину $g\%$ (от максимального значения $\ln 2$). Коэффициент « $-g$ » в этом выражении определяет степень изменения энтропии с изменением вероятности состояния или, иначе, характеризует интенсивность изменения энтропии.

Далее удобно обратиться к геометрической иллюстрации энтропии H (рис. 3в). Этот рисунок можно рассматривать как дополнение к рис. 3а и 3б.

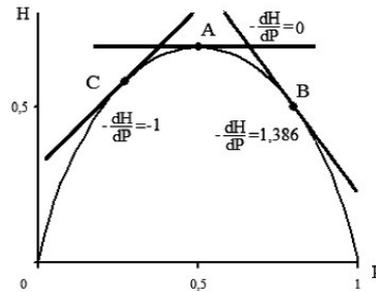


Рис. 3в. Информационная трактовка модели Раша: средняя энтропия H

При $P = 0,5$ (точка А) энтропия H достигает максимума, касательная к графику функции принимает горизонтальное положение, потеря информации нет, аналитически производная равна нулю $\frac{dH}{dP} = 0$.

В этом состоянии уровни подготовленности испытуемого и тестового задания совпадают, $g = b - d = 0$.

С изменением значения P , возрастания или убывания, эн-

тропия начинает уменьшаться, т.е. происходит потеря информации. Чем дальше отклонение P от значения 0,5, тем интенсивней происходит потеря информации, причём интенсивность этих потерь как раз и определяет значение показателя

$-\frac{dH}{dP}$ (знак минус отражает

уменьшение информации с ростом P). Таким образом, информационный смысл показателя

$g = -\frac{dH}{dP}$ состоит в отражении

интенсивности информационных потерь.

Так, в точке В касательная наклонена к графику по углом

$tg = \frac{dH}{dP} = -1,386$, энтропия в

этой точке убывает со скоростью 1,386 нат, что соответствует значениям показателей $g = b - d = 1,386$. В точке С касательная наклонена к графику под углом 45° , на каждый процент роста вероятности правильного ответа количество информации H увеличивается на 1%. Значения традиционных показателей IRT этого состояния $g = b - d = -1$.

Отсюда можно установить: показатель g имеет ещё одну информационную трактовку — интенсивность информацион-

ных потерь $-\frac{dH}{dP}$.

К группе основных показателей IRT можно отнести и информационную функцию $I = P \cdot (1 - P)$, безразмерные значения которой служат индикатором качества измерений, иногда мерой информации о локальном измерении. Это частнонаучный показатель. Его общенаучным аналогом является информационная энтропия Шеннона H с информационными единицами измерения. В этом её основное достоинство.

Вторым достоинством информационной функции $H(x)$ является её универсальность — независимость от функции распределения вероятности P события x . Информационная функция I этим свойством универсальности не обладает. Информационная функция напрямую связана с дисперсией распределения вероятности, способы и возможность вычисления которой определяются функцией распределения.

Таким образом, к традиционной системе основных показателей IRT, выраженного в логитах $g = b - d$, и безразмерного значения информационной функции I , можно поставить в соответствие информационные показатели, измеряемые в единицах количества информации — разности частных энтропий противоположных событий h_{01} или интенсивности информационных потерь

$-\frac{dH}{dP}$ и меры неопределённости

исходного состояния H .

Конструирование модели педагогических измерений

Процедура вывода математической модели Раша из формулы Шеннона позволяет предложить алгоритм конструирования целого класса математических моделей педагогических измерений. Этот алгоритм достаточно прост и кратко может быть выражен двумя этапами.

Первый — по заданному виду тестового задания сформировать функции Шеннона для расчёта заложенной меры неопределённости в данном задании.

Второй этап — используя связи традиционных показателей с информационными, выразить искомую вероятность правильного ответа P , т.е. построить модель педагогических измерений.

В качестве первого примера сконструируем хорошо известную двухпараметрическую модель А. Бирнбаума.

Идея Бирнбаума состоит во введении в модель Раша дополнительного параметра для характеристики дифференцирующей способности задания. При конструировании такой модели достаточно в информа-

Методология

ПЕД
измерения

ционно-измерительное соответствие $h_{01} \Leftrightarrow g$ ввести множитель a дифференцирующей способности задания: $h_{01} = a \cdot g$. Этот множитель увеличивает (при $a > 1$) или уменьшает (при $0 < a < 1$) «чувствительность» латентной разности g на наличие разности частных неопределённостей h_{01} . Тогда можно

записать $\ln \frac{P}{1-P} = a \cdot g$. Отсю-

да, выражая вероятность правильного ответа P , имеем хорошо известную двухпараметрическую модель Бирнбаума:

$$P = \frac{\exp(a(b-d))}{1 + \exp(a(b-d))}$$

Двухпараметрическую модель Бирнбаума можно схематически представить ориентированным графом (рис. 4). Определим его. Вершины такого графа обозначают возможные состояния выполнения тестового задания – соответствующие вероятности P с началом в вершине «неправильный ответ», дуга графа – направленная линия с потенциалом, равным разности частных энтропий смежных вершин (обозначено над дугой графа). Потенциалу дуге графа соответствует свой измеряемый показатель IRT (обозначено под дугой графа).

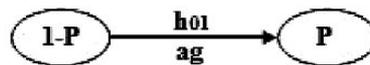


Рис. 4. Граф модели А.Бирнбаума

Второй пример. Сконструируем педагогическую модель для тестового задания с тремя градациями ответа: неправильный – частично правильный – правильный. Обозначим вероятность частично правильного ответа через P_1 (1 балл), правильного ответа P_2 (2 балла), тогда вероятность неправильного ответа $P_0 = 1 - P_1 - P_2$. Разницы частных энтропий и средняя энтропия рассматриваемой системы определяются выражениями:

$$h_{01} = h_0 - h_1 = -\ln(1 - P_1 - P_2) + \ln P_1,$$

$$h_{12} = h_1 - h_2 = -\ln P_1 + \ln P_2,$$

$$H(x) = -(1 - P_1 - P_2)$$

$$\cdot \ln(1 - P_1 - P_2) - P_1 \cdot \ln P_1 - P_2 \cdot \ln P_2.$$

В этом примере продолжим построение модели в двух вариантах: через определение разности частных энтропий и через вычисление производной средней энтропии.

Для демонстрации первого варианта удобно построить граф, отражающий логику выполнения тестового задания (рис. 5а). Условие тестового задания ($У$) предполагает два варианта его выполнения (B_1 и B_2): с частично правильным ответом на поставленный в задании вопрос и уточняющий полный ответ. Т.е. предполагается, что полный ответ не независим, а следует после

предварительно правильного ответа.

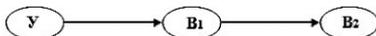


Рис. 5а. Граф задания с частично правильным ответом

Процесс выполнения первой части задания можно описать уже привычной формулой с выделением индекса у показателя g для первого задания $h_{01} = g_1$. Процесс выполнения всего задания можно представить как сумму последовательных этапов его выполнения: $h_{01} + h_{12} = g_1 + g_2$. Разрешая систему этих уравнений относительно P_1 и P_2 , находим искомые выражения для расчётов вероятности правильного ответа, которые являются двухшаговым вариантом известной модели РСМ (Partial Credit Model):

$$P_1 = \frac{\exp(g_1)}{1 + \exp(g_1) + \exp(g_1 + g_2)};$$

$$P_2 = \frac{\exp(g_1 + g_2)}{1 + \exp(g_1) + \exp(g_1 + g_2)}.$$

По структуре графа тестового задания несложно привести и граф модели педагогических измерений для этого задания (рис. 5б).

Второй вариант построения модели отличается только формой математических преобразований. Т.к. функция H зависит от двух переменных, то находим частные производные

и приравниваем их к соответствующим показателям педагогических измерений:

$$-\frac{\partial H}{\partial P_1} = \ln \frac{P_1}{1 - P_1 - P_2} = g_1;$$

$$-\frac{\partial H}{\partial P_2} = \ln \frac{P_2}{1 - P_1 - P_2} = g_1 + g_2.$$

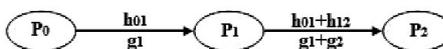


Рис. 5б. Граф двухшаговой модели РСМ

Разрешая систему уравнений относительно P_1 и P_2 , приходим к полученной выше модели.

В качестве третьего примера разработаем модель для задачи по физике.

В закрытом сосуде объёма V находится идеальный газ в количестве n молей. Определить: 1. Температуру газа T , при которой давление в сосуде будет равно атмосферному; 2. На сколько увеличится давление газа при его нагреве на ΔT градусов.

Критерии оценки. Полное и правильное решение всей задачи оценивается в 3 балла. За правильное выполнение каждого из трёх пунктов назначается 1 балл: а) правильно записано уравнение состояния идеального газа; б) записаны правильные решение и ответ на вопрос задачи 1; в) записаны правильные решение и ответ на вопрос задачи 2.

В тексте задания и критериальных требованиях можно вы-

ПЕД
измерения

делить условие задачи (У) и три вопроса (В1, В2, В3), по которым можно построить граф тестового задания (рис. 6а). При построении графа учтено, что правильное выполнение пунктов б) и в) основывается на правильном выполнении пункта а), но между собой решения б) и в) независимы.

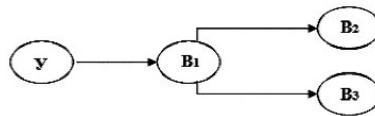


Рис. 6а. Граф задания

Конструирование модели

Обозначим вероятность правильного выполнения пункта а) через P_1 , пункта б) — через P_2 , пункта в) — через P_3 , тогда вероятность неправильного ответа равна $P_0 = 1 - P_1 - P_2 - P_3$. Полная энтропия определяется выражением:

$$H(x) = -(1 - P_1 - P_2 - P_3)$$

$$\cdot \ln(1 - P_1 - P_2 - P) - P_1 \cdot \ln P_1 -$$

$$- P_2 \cdot \ln P_2 - P_3 \cdot \ln P_3.$$

Находим частные производные функции H и приравниваем их к показателям, соответствующим информационным потерям:

$$-\frac{\partial H}{\partial P_1} = \ln \frac{P_1}{1 - P_1 - P_2 - P_3} = g_1;$$

$$-\frac{\partial H}{\partial P_2} = \ln \frac{P_2}{1 - P_1 - P_2 - P_3} = g_1 + g_2;$$

$$-\frac{\partial H}{\partial P_3} = \ln \frac{P_3}{1 - P_1 - P_2 - P_3} = g_1 + g_3.$$

Разрешая систему уравнений относительно P_1 , P_2 и P_3 , находим искомые модели:

$$P_1 = \frac{\exp(g_1)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_3)};$$

$$P_2 = \frac{\exp(g_1 + g_2)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_3)};$$

$$P_3 = \frac{\exp(g_1 + g_3)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_3)}.$$

Конструирование модели на основе частных энтропий h приводит к той же модели и здесь не приводится. Граф модели представлен на рис. 6б:

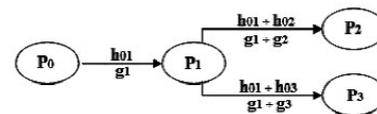


Рис. 6б. Граф модели педагогических измерений для тестового задания

Рассмотренные примеры позволяют рекомендовать при конструировании моделей с разветвлённой структурой тестовых заданий использовать

графы. С учётом этого можно предложить следующий упрощённый пошаговый алгоритм конструирования измерительной модели тестового задания.

Первый шаг — построение графов. По заданному условию тестового задания и критерияльным требованиям к оцениваемым элементам ввести необходимые обозначения и построить граф тестового задания. Не изменяя структуры этого графа, ввести дополнительные обозначения и построить изоморфный граф модели, служащий основой для построения математической модели измерений для данного тестового задания (пример: рис. 6а и 6б к задаче по физике).

Второй шаг — построение модели педагогических измерений. Вероятность правильного ответа P_i на i -й элемент контроля тестового задания для аддитивной модели класса модели Раша определяется по формуле:

$$P_i = \frac{\text{работа на пути от } P_0 \text{ до } P_i}{\text{работа на графе}},$$

где: путём на графе от начальной вершины P_0 до выбранной вершины P_i будем называть последовательность дуг графа, соединяющих эти вершины; потенциалом пути будем называть разность частных энтропий начальной и выбранной вершин пути h_{0i} ; работой на

этом пути будем называть экспоненту от суммы измеряемых показателей, соответствующих потенциалу пути $\exp(g_1 + g_2 + \dots + g_i)$. Работа на графе равна сумме работ на всех возможных путях графа, включая работу на его единственной петле от P_0 до P_0 равную $\exp(0) = 1$.

Например, путь от P_0 до P_3 через вершины P_1 и P_2 представляет собой последовательность дуг с суммарным потенциалом $h_{03} = h_{01} + h_{12} + h_{23}$, работа на этом пути определяется выражением $\exp(g_1 + g_2 + g_3)$. Если в основе графа лежат только эти четыре вершины, то граф содержит 4 различных пути — $P_0 - P_0$, $P_0 - P_1$, $P_0 - P_1 - P_2$, $P_0 - P_1 - P_2 - P_3$, с суммарной работой на графе $1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)$.

Таким образом, описанный процесс конструирования модели педагогических измерений даёт разработчику теста свободу в выборе формы тестовых заданий. Теперь не форма тестового задания подбирается под известные модели педагогических измерений, а наоборот, на заданную форму тестового задания может быть сконструирован свой закон расчёта вероятностей правильного ответа на его задания.

ПЕД измерения

Модель измерения педагогических отношений

В предыдущих работах мы обозначали эту модель как альтернативную, вкладывая в определение «альтернативная» смысл параллельности к аддитивной модели Раша и её модификациям. Отличительной особенностью этой модели является возможность проведения педагогических измерений в пропорциональной шкале, т.е. возможность проведения измерений педагогических отношений — относительных уровней подготовленности испытуемых и и относительных уровней трудности тестовых заданий β . Сама математическая модель имеет вид: $P = 0,5^{\beta/\theta}$.

Применение информационного подхода к описанию модели измерения педагогических отношений принимает ещё более простой вид. Если из выражения для частной энтропии $h = -\log_2 P$ выразить вероятность P , то получим общий вид альтернативной модели педагогических измерений $P = 0,5^h$. И вопрос построения модели фактически превращается в вопрос об интерпретации показателя h применительно к проблеме педагогических измерений.

Вариант такой интерпретации представим с иллюстрациями на рис. 7. Монотонно убывающая функция частной энтропии $h(P)$ определена на интервале $(0; 1]$ с областью значений $(\infty; 0]$. Для достоверного события $P = 1$, неопределённость такого состояния $h(1)=0$ бит. Применительно к тестовым испытаниям для участника с уровнем подготовленности θ это событие эквивалентно нулевому уровню трудности тестового задания $\beta = 0$. С ростом трудности тестового задания вероятность правильного ответа испытуемым убывает, а неопределённость такого состояния растёт. Т.е. с ростом уровня трудности задания β при постоянном уровне θ растёт неопределённость состояния h и вполне допустимо предположить эту связь пропорциональной $h \sim \beta$.

вающая функция частной энтропии $h(P)$ определена на интервале $(0; 1]$ с областью значений $(\infty; 0]$. Для достоверного события $P = 1$, неопределённость такого состояния $h(1)=0$ бит. Применительно к тестовым испытаниям для участника с уровнем подготовленности θ это событие эквивалентно нулевому уровню трудности тестового задания $\beta = 0$. С ростом трудности тестового задания вероятность правильного ответа испытуемым убывает, а неопределённость такого состояния растёт. Т.е. с ростом уровня трудности задания β при постоянном уровне θ растёт неопределённость состояния h и вполне допустимо предположить эту связь пропорциональной $h \sim \beta$.

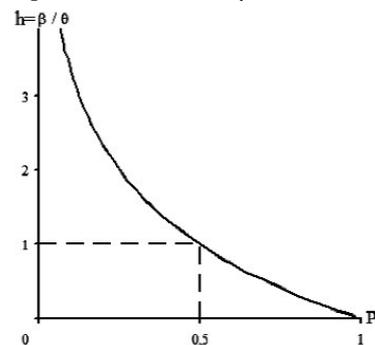


Рис. 7. График функции частной энтропии $h(P)$

Аналогичные рассуждения можно привести и для фиксированного задания с уровнем трудности β . Тогда, достоверно правильное решение задания

можно объяснить бесконечно высоким уровнем подготовленности испытуемого. Падение уровня подготовленности испытуемого и приводит к падению вероятности правильного ответа P и росту неопределённости состояния h , т.е. опять вполне логично предположить обратную пропорциональную связь $h \sim 1/\theta$.

Коэффициент пропорциональности легко находится из следующих соображений: когда уровни подготовленности испытуемого и трудности тестового задания совпадают $\beta = \theta$ или $\beta/\theta = 1$, вероятность правильного ответа на задание равна 0,5 и мера неопределённости такого состояния определяется значением $h = 1$ бит. Отсюда имеем $h = \beta/\theta$ или окончательно выражение модели измерения педагогических отношений: $P = 0,5\beta/\theta$.

Таким образом, в альтернативной модели отношение показателей β/θ есть не что иное, как мера неопределённости состояния «правильный ответ испытуемого на тестовое задание», выраженное значением частной энтропии Шеннона h .

В отличие от информационного толкования моделей педагогических измерений класса Раша, где измеряемые параметры подготовленности и трудности выражаются через разность частных энтропий состояний «неправильный ответ» и «пра-

вильный ответ», измеряемые параметры альтернативной модели выражаются только через частную энтропию состояния «правильный ответ». Такая зависимость только от одного состояния позволяет использовать альтернативную модель измерений автономно, без каких либо процедур согласования, не только к дихотомическим тестовым заданиям, но и к заданиям с произвольным уровнем градации ответов. На наш взгляд, единственно возможным «совершенствованием» такой модели может быть введение множителя дифференцирующей способности задания в виде степени к показателю уровня трудности тестового задания β^a .

Критерии эффективности теста

Аналитические свойства информационной функции Шеннона $H(x)$ позволяют использовать её при разработке критериев эффективности теста. Существование максимума энтропии можно определить как оптимальное информационное состояние вероятностной системы. Тогда отношение энтропии наблюдаемого состояния системы к энтропии оптимального состояния может служить количественной оценкой информационной эффективности

ПЕД
измерения

ти. Эти достаточно общие рассуждения могут найти применения и в задачах разработки критериев эффективности педагогических тестов.

Рассмотрим две таких задачи — задача оценки эффективности информационного содержания теста и задача оценки эффективности ранжирования участников теста.

В первой задаче в качестве наблюдаемого критерия эффективности можно принять среднее значение энтропии всей системы вероятностей правильного ответа испытуемых на тестовые задания. Такие вычисления допустимы свойством аддитивности функции Шеннона. Если вероятность правильного ответа i -м испытуемым на j -е тестовое задание обозначить через P_{ij} , то относительное значение (в расчёте на один локальный акт ответа одним испытуемым на одно задание теста) информационной энтропии теста будет определяться по формуле:

$$H_{\text{омн.}} = \frac{1}{N \cdot M}$$

$$\cdot \sum_i \sum_j \left[-P_{ij} \cdot \log_2 P_{ij} - (1 - P_{ij}) \log_2 (1 - P_{ij}) \right]$$

где N — количество испытуемых, M — количество тестовых заданий.

Применение этого критерия следует сопроводить несколькими замечаниями.

Первое. Максимально возможное значение этой функции $H_{\text{max}} = 1$ бит. Однако в практике это состояние недостижимо. Это тот невероятный случай, когда уровни подготовленности всех испытуемых совпадают с уровнями трудности всех тестовых заданий и вероятность каждого правильного ответа равна 0,5.

При равномерном распределении вероятности правильного ответа значение функции Шеннона $H_{\text{омн.}}$ равно $1/2 \ln 2$. Тогда эффективность такого состояния оценивается значением

$$\frac{H_{\text{омн.}}}{H_{\text{max}}} = \frac{1}{2 \cdot \ln 2} \approx 0,72$$

или, в процентном выражении, 72%. Для реального теста такое значение показателя эффективности информационного содержания можно рассматривать как допустимое.

На наш взгляд, принимать этот критерий как целевой при разработке тестовых заданий правомерно лишь для решения задач максимального сбора информации о подготовленности всей группы испытуемых. Например, этот критерий эффективности теста можно использовать при контроле знаний группы обучающихся в процессе обучения.

Дело в том, что среднее значение информационной энтропии теста растёт с увеличе-

нием доли заданий со средним уровнем трудности. И критерий $H_{отн.}$ принимает значения, близкие к максимуму, когда все задания имеют уровень трудности, близкий к среднему. Но такие тесты имеют и недостатки. С метрической точки зрения, такие тесты выводят результаты измерений слабо и сильно подготовленных испытуемых в область высоких погрешностей, вплоть до исключения из анализа результатов теста. Дидактические недостатки таких тестов состоят в том, что сама природа учебных знаний имеет различные уровни глубины сложности содержания и, соответственно, трудности их усвоения. Приводить все контрольные задания к золотой середине, выявлять только средние знания и нецелесообразно, и не представляется возможным.

Второе замечание относится к технологии обработки тестовых результатов. Дело в том, что значение информационной функции теста H зависит от принятой математической модели педагогических измерений. И может сложиться впечатление, что этот критерий можно использовать при выборе оптимально-эффективной модели педагогических измерений и определении её параметров. Назначение этого критерия другое. Этот критерий отражает ин-

формационно-содержательную эффективность подбора тестовых заданий для группы испытуемых. Это свойство не должно зависеть от метода обработки результатов теста.

Рассмотрим задачу оценки эффективности ранжирования тестируемых испытуемых. Проблема ранжирования испытуемых обычно ставится перед тестом при проведении конкурсных отборов, когда основной целью тестовых испытаний является максимальное разъединение испытуемых по уровням подготовленности относительно содержания теста. Приведём эту задачу к математическому виду.

Введём индекс ранга k по числу правильных ответов и обозначим через n_k количество испытуемых, правильно ответивших на k тестовых заданий. Тогда для N испытуемых справедливо равенство $N = \sum_k n_k$ и можно ввести понятие доли испытуемых, принадлежащих k -му рангу $\omega_k = \frac{n_k}{N}$,

с условием нормировки $\sum \omega_k = 1$. В этих обозначениях функцию Шеннона для информационной энтропии можно записать в виде: $H_{ранж.} = -\sum \omega_k \cdot \log_2 \omega_k$,

где суммирование происходит по всем значения k от 1 до $M - 1$. Последнее выражение можно трактовать как энтро-

ПЕД
измерения

пия сообщения о ранжировании испытуемых в результате тестовых испытаний по всем возможным ранговым группам. Максимально возможное значение $H_{\max} = \log_2(M - 1)$ эта функция принимает при равномерном распределении испытуемых по ранговым группам. Тогда эффективность ранжирования участников теста можно характеризовать показателем $H_{\text{ранж}}/H_{\max}$.

Совершенно аналогично можно ввести и показатель эффективности ранжирования тестовых заданий по уровням трудности. Для хорошо ранжированного теста оба показателя эффективности должны быть близки к единице. Примером идеального ранжирования испытуемых и заданий может служить идеалистический тест Л. Гутмана, когда профили подготовленности всех испытуемых правильные и каждое задание чётко различает одного испытуемого.

Примеры расчётов показателей эффективности теста

Для иллюстрации свойств введённых показателей проведём расчёты для трёх модельных таблиц тестовых данных, представленных на рис. 8. Таблицы содержат результаты ответов 8 испытуемых на 5 тестовых заданий. В правом столбце таблиц приведены доли правильных ответов каждым испытуемым, в нижней строке — доли правильных ответов на отдельное тестовое задание. Общая доля правильных ответов во всех трёх таблицах данных равна 0,5.

Результаты расчётов представлены в таблице. При расчёте показателей эффективности информационного содержания мы использовали значения вероятности правильного ответа P_{ij} , полученные на основе исходных решений схемы Раша. Если данные таблиц а) и б) дают значения близкие к допус-

1	1	1	1	0	0,8	1	1	1	1	0	0,8	1	1	0	1	0	0,6
1	1	1	0	1	0,8	1	1	1	0	1	0,8	1	1	1	0	0	0,6
1	1	1	0	0	0,6	1	1	1	0	0	0,6	1	1	1	0	0	0,6
1	1	0	1	0	0,6	1	1	1	0	0	0,6	1	1	1	0	0	0,6
1	1	0	0	0	0,4	1	0	0	1	0	0,4	0	0	0	1	1	0,4
1	0	1	0	0	0,4	1	0	0	1	0	0,4	0	0	0	1	1	0,4
1	0	0	0	0	0,2	1	0	0	0	0	0,2	1	0	0	0	1	0,4
0	1	0	0	0	0,2	0	0	0	1	0	0,2	0	0	1	1	0	0,4
0,875	0,75	0,5	0,25	0,125	0,5	0,875	0,5	0,5	0,5	0,125	0,5	0,625	0,5	0,5	0,5	0,375	0,5
а)						б)						в)					

Рис. 8. Таблицы результатов тестовых испытаний

Таблица

Методология

Информационные показатели эффективности теста

Таблица данных	Показатель эффективности информационного содержания	Показатель эффективности ранжирования		
		испытуемых	заданий	теста
а)	0,69	1	0,83	0,83
б)	0,73	1	0,49	0,49
в)	0,95	0,5	0,49	0,24

тимому уровню 0,72, то данные третьей таблицы дают очень высокий результат — эффективность информационного содержания теста составляет 95%. Этот результат — предсказуем. Именно в третьей таблице доли правильных ответов, как испытуемых, так и на тестовые задания, сгруппированы около среднего значения 0,5, а значит, и вероятности правильных ответов на тестовые задания сгруппированы около информационно оптимального значения 0,5.

Проанализируем данные трёх таблиц с позиций эффективности ранжирования испытуемых и тестовых заданий. Сначала можно привести несколько общих замечаний для всех таблиц: 5 тестовых заданий позволяют выделить 4 ранговые группы для ранжирования 8 испытуемых; 8 испытуемых позволяют выделить семь ранговых групп для ранжирования тестовых заданий.

Рассмотрим, как в приведённых примерах распределены

испытуемых 5 тестовых заданий. В таблицах *а)* и *б)* результаты испытуемых равномерно распределены в четыре ранговые группы по два участника теста в каждой. Каждая ранговая группа одинаково заполнена испытуемыми, эффективность ранжирования испытуемых оптимальная с единичным значением показателя. В таблице *в)* результаты испытуемых сгруппированы в две ранговые группы по четыре участника в каждой, а две ранговые группы не заполнены. В этом случае набор тестовых заданий не разделил участников теста по всем возможным ранговым группам, показатель эффективности ранжирования отразил этот результат значением 0,5.

Рассмотрим распределения заданий. При ранжировании 5 тестовых заданий по 7 возможным уровням неизбежны вакансии. Но если в таблице *а)* только две ранговые группы оказались свободными, в таблицах *б)* и *в)* тестовые задания

ПЕД	
	измерения

распределены в три группы в количествах 1, 3, 1, оставив незаполненными 4 ранговые группы. Эти свойства теста отражены значениями показателей эффективности тестовых заданий 0,83 для таблицы данных *а)* и 0,49 для данных таблиц *б)* и *в)*. И наконец, в последнем столбце таблицы приведены показатели эффективности ранжирования всего теста, как произведения показате-

лей эффективностей ранжирования испытуемых и тестовых заданий.

Таким образом, введённые информационные показатели эффективности количественно отражают выделенные свойства теста и могут найти применение как в анализе результатов тестирования, так и в виде рекомендаций к конструированию нового теста или совершенствованию имеющегося.