

УДК 378.091.279.1:004.9

Л. С. Мороз, ассистент (БГТУ); Е. В. Кабак, ассистент (БГТУ);
А. И. Бракович, ст. преподаватель (БГТУ)

СРАВНИТЕЛЬНАЯ ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ КОМПЬЮТЕРНОГО КОНТРОЛЯ ЗНАНИЙ

Статья посвящена обзору существующих моделей и методов оценки знаний в процессе компьютерного тестирования. Материал статьи включает классификацию типов тестовых заданий, которая учитывает мнения разных исследователей, работающих в данной области знаний, методы количественной оценки знаний студентов по каждому из перечисленных типов заданий. Особое внимание авторами уделяется проблеме учета в процессе формирования оценки по результатам тестирования различных параметров, как отдельных тестовых заданий, так и всего компьютерного теста в целом: время тестирования, вероятность угадывания правильного ответа тестируемым, тип тестового задания, уровень сложности задания и т. д. Данная статья будет полезна специалистам, занимающимся проблемами организации и осуществления компьютерного контроля знаний, разработки специального программного обеспечения, обеспечивающего поддержку названных процессов.

Article is devoted to the review of existing models and methods of knowledge estimation in the computer testing process. Article material includes types (forms) classification of test tasks, uniting opinions of various researchers in the given scientific field, methods of a quantitative estimation of students' knowledge on each of the listed tasks types. The special attention in this article is given by authors to a problem of the account in the process of estimation formation by testing results of various test parameters, both separate test tasks, and the computer test as a whole: testing time, probability of guessing of a right answer, type of the test task, complexity level of the task etc. Given article will be useful for the experts (teachers, scientists, software developers, ect.), dealing with problems of the organization and realization of the computer control of knowledge, development of the special software providing support of named process.

Введение. Контроль знаний – действенный стимул учебного процесса, его важная составная часть, позволяющая получить всестороннюю оценку уровня знаний обучаемых [1].

Одной из популярных и хорошо зарекомендовавших себя форм контроля знаний является компьютерный контроль знаний, который реализуется за счет специального программного обеспечения.

Компьютерные тесты – одна из наиболее широко распространенных форм компьютерного контроля знаний, которая представляет собой высокотехнологическое средство для измерения уровня знаний обучаемых, обеспечивающее возможность его своевременной коррекции, а также изменения методики подготовки и проведения занятий.

Проблема проведения компьютерного контроля знаний (компьютерного тестирования) является весьма сложной и противоречивой, что обусловлено целым рядом факторов и параметров, требующих обязательного учета. Особую важность здесь приобретает проблема объективной количественной оценки знаний обучаемых, предполагающая правильный выбор методов оценивания знаний. В настоящее время этой проблемой занимаются исследователи из различных областей знаний: ученые, педагоги, специалисты в области информационных технологий и др.

Целью данной статьи является рассмотрение и анализ современных методов оценки знаний обучаемых в процессе компьютерного тестирования.

Основная часть. Многие исследователи сходятся в том, что в процессе количественной оценки знаний необходимо учитывать такие параметры, как уровень сложности тестовых заданий, тип (форму) заданий, вероятность угадывания правильных вариантов ответов обучаемыми, и другие факторы, позволяющие, например, индивидуализировать процесс контроля знаний, адаптировать учебный тест к каждому конкретному испытуемому. Так, например, разделение тестовых заданий по уровням сложности обеспечивает достаточную объективность оценивания знаний, дает возможность получить представление об истинном уровне знаний обучаемых. Использование заданий различных типов позволяет сформировать тест, отличный от простого угадывания правильных ответов, при этом некоторые из заданий (на установление соответствия, на установление правильной последовательности) дают возможность уменьшить количество неверной информации на экране, отражают совокупность взаимосвязанных понятий.

В современных условиях компьютерный контроль знаний может быть осуществлен с использованием различных методов формирования

оценки. Мы будем осуществлять их отбор с учетом простоты технической реализации, возможности оценивания с их помощью знаний в зависимости от типа тестового задания, подходов к формированию итоговой оценки в целом по тесту (нескольким тестам) или по отдельным тестовым заданиям.

В самом простом случае итоговая оценка может формироваться на основании оценок, полученных за отдельные вопросы (тестовые задания).

В работе И. П. Карповой [2] приводятся следующие методы формирования итоговой оценки.

1. «Сумма» – итоговая оценка O равна сумме баллов O_i , полученных за отдельные i -е ответы:

$$O = \sum_{i=1}^k O_i. \quad (1)$$

2. «Шкала» – итоговая оценка O подсчитывается в соответствии со шкалой, заданной минимальным \min и максимальным \max баллами:

$$O = \frac{\sum_{i=1}^{N_q} O_i (\max - \min) + \min}{N_q}, \quad (2)$$

где O_i – оценки, полученные за отдельные ответы; N_q – общее число заданных вопросов.

3. «Процент» – итоговая оценка O подсчитывается как процент правильных ответов:

$$O = \frac{m}{N_q} 100\%, \quad (3)$$

где m – число правильных ответов; N_q – общее число заданных вопросов.

Однако более интересной является предложенная в [3] дифференциальная модель процесса адаптивного тестирования знаний, иначе описывающая процедуру формирования результата выполнения некоторого теста R . В соответствии с этой моделью результат выполнения теста изменяется непрерывно и дифференцируем как функция, зависящая от количества выполненных заданий. На результат прохождения тестирования, согласно этой модели, влияют:

- уровень сложности заданий z_i ($z_i = \overline{1, Z}$);
- шкала оценивания (рекомендуется использовать непрерывную шкалу оценивания: за выполнение каждого тестового задания тестируемый получает некоторый коэффициент $r_i \in [0; 1]$, определяемый индивидуально для каждой формы тестовых заданий);
- вероятность угадывания правильных вариантов ответа q_i ;
- балльность системы B .

При учете введенных параметров, влияющих на результат прохождения теста, динамику изменения результата выполнения теста при выполнении каждого задания в [3] предложено описывать формулой

$$\frac{dR}{dN} = \left(r_i z_i + \frac{dR}{d(N-1)} \frac{(1-q_i)(\dot{Z} - z_i)}{B} \right) \frac{B}{\dot{Z}}. \quad (4)$$

Здесь $\frac{dR}{dN}$ – дифференциальное оценивание результата прохождения теста в данный момент времени; r_i – коэффициент оценивания заданий разных типов, изменяемый в диапазоне $[0; 1]$; z_i – уровень сложности i -го задания; $\frac{dR}{d(N-1)}$ –

дифференциальное оценивание результата прохождения теста в предыдущий момент времени (при выполнении $N-1$ задания); N – количество тестовых заданий, выполненных в данный момент времени; q_i – вероятность угадывания (ввода) правильного ответа; B – балльность

системы; $\dot{Z} = \sum_{i=1}^{N_t} z_i$ – суммарная сложность теста, где i – номер задания, а N_t – номер текущего выполняемого задания.

Как мы уже неоднократно отмечали, создание эффективного теста предполагает включение в него тестовых заданий различного типа. В настоящее время существует несколько классификаций типов тестовых заданий. Мы будем в основном придерживаться классификации, которая приведена в работе [4], дополненной и расширенной в соответствии с классификациями, выдвинутыми в работах [1–3]. Предлагаемая классификация типов тестовых заданий состоит из заданий:

- закрытого типа (с множественным выбором (многоальтернативные), с одиночным выбором (одноальтернативные), с градуированными ответами), в которых тестируемый выбирает правильный ответ из данного набора ответов;
- открытого типа (различные задания на дополнение, например на введение термина или определения, на заполнение таблиц и др., а также задания свободного изложения), требующие от тестируемого самостоятельно сформулировать ответ;
- на установление соответствия, выполнение которых связано с выявлением соответствия между элементами двух множеств;
- на установление правильной последовательности, в которых тестируемый должен указать верный порядок действий или процессов;

– многошаговых (ситуационных), которые состоят из некоторого ограниченного набора шагов (заданий), решаемых последовательно.

Следует сказать несколько слов о заданиях закрытого типа с градуированными ответами. Задания с *градуированными ответами* (второе название – задания с выбором наилучшего ответа) содержат ответы, из которых все, возможно, являются правильными в той или иной степени. Ответы имеют градацию по степени правильности. Задача составителя заключается в том, чтобы найти и применить признак, позволяющий осуществить такую градацию. Максимальное количество баллов испытуемый получает, если его градация ответов полностью совпадает с градацией эксперта, например разработчика задания.

Для объективной оценки тестовых заданий разных типов предлагается использовать для каждого из них свой специализированный подход для формирования оценки.

Классическая теория тестов основана на разложении наблюдаемого множества ответов на множества верных и неверных ответов.

Для оценивания *закрытого тестового задания с одиночным выбором* широко используется дихотомическая шкала оценивания, в которой за каждое задание можно получить 0 или 1 балл, т. е. 1 соответствует правильному ответу, а 0 – неправильному. Если обозначить коэффициент оценивания для ответов на тестовые задания разных форм в отдельности через r_i , как это сделано в работах [1, 3], получим:

$$r_i = \begin{cases} 1, & \text{если } i\text{-й ответ верный,} \\ 0, & \text{если } i\text{-й ответ неверный.} \end{cases} \quad (5)$$

При этом вероятность случайного ввода правильного ответа определяется следующей формулой [1–3]:

$$q_1 = \frac{1}{N}, \quad (6)$$

где N – число всех вариантов ответа в задании.

Задания с одиночным выбором справедливо критикуются за довольно высокую вероятность угадывания верного ответа.

Этого недостатка лишены задания с множественным выбором, в блоке ответов которых размещено несколько верных ответов (*дистракторов*). Однако повышение количества верных ответов в заданиях указанного типа приводит к необходимости общего увеличения числа ответов в блоке ответов. Если считать оптимальным соотношение один верный ответ на два дистрактора, то при трех верных ответах потребуется уже 6 дистракторов, итого 9 вариантов ответов.

Кроме того, оценивание результата выполнения такого задания испытуемым сложнее, чем оценивание задания с выбором одного верного ответа.

Для оценивания *закрытого тестового задания с множественным выбором* коэффициент r_i предлагается рассчитывать по формуле

$$r_i = \frac{Q_2}{Q_1 + Q_3}, \quad (7)$$

где Q_2 – количество правильных вариантов ответа, выбранных обучаемым; Q_1 – количество всех правильных вариантов ответа в задании; Q_3 – количество неправильных вариантов ответа, выбранных обучаемым.

Вероятность случайного ввода правильного ответа для заданий с множественным выбором может определяться следующим соотношением:

$$q_2 = \frac{1}{M}, \quad (8)$$

где M – общее количество возможных ответов и $M = \sum_{k=0}^N C_N^k = \sum_{k=0}^N \frac{N!}{k!(N-k)!}$, здесь N – число всех вариантов ответа, из них k – правильных.

Поскольку $\sum_{k=0}^N C_N^k = 2^N$, то $q_2 = \frac{1}{2^N}$.

Как видно из приведенных формул, преимуществом заданий с множественным выбором по сравнению с заданиями с одиночным выбором является более высокая защищенность от угадывания. Однако весьма непросто создать задание, содержащее и несколько верных ответов и большое количество сильных дистракторов. При слабых дистракторах защищенность от угадывания будет сильно снижаться.

Серьезным недостатком заданий с одиночным ответом и с множественными ответами является тот факт, что испытуемые могут запомнить неверные ответы (дистракторы), и это приведет к снижению их уровня подготовленности. В заданиях с градуированными ответами можно дать все ответы верные, но в различной степени, что сильно усложнит задачу испытуемого. Сложность таких заданий обусловлена тем, что испытуемому надо не просто выбрать верный ответ, но еще дать свою оценку остальным ответам. Угадывание в такой ситуации практически исключено.

При оценке тестового задания на установление соответствия каждую пару ответов рассматривают как отдельный вариант ответа и при выставлении результата выполнения задания учитывают, сколько пар было выбрано верно. В этом случае коэффициент r_i следует рассчитывать по следующей формуле:

$$r_i = \frac{Q_2}{Q_1}, \quad (9)$$

где Q_2 – количество правильно составленных пар; Q_1 – количество пар для сопоставления.

В случае тестового задания на соответствие вероятность случайного ввода правильного ответа может высчитываться разными способами в зависимости от способов представления ответов.

Первый способ предполагает использование двух списков по m элементов, из которых необходимо составить m пар.

В этом случае расчет может быть осуществлен по формуле

$$q_3 = \frac{1}{(m!)^2}. \quad (10)$$

Второй способ предполагает использование двух списков по m и n элементов, причем, например, $m \leq n$. Тогда, согласно выражению

$$q_4 = \frac{(n-m)!}{n!m!}. \quad (11)$$

Третий способ рассматривает случай, когда имеется два списка по m и n элементов, причем $m \leq n$ и необходимо составить l пар, где $l < m$.

Тогда расчет может быть выполнен по следующей формуле:

$$q_5 = \frac{(m-l)!(n-l)!}{m!n!}. \quad (12)$$

При выполнении тестового задания на установление правильной последовательности, как и в случае одноальтернативного тестового задания, возможен только один заведомо правильный ответ, поэтому для оценивания данного типа тестовых заданий также достаточно использовать дихотомическую шкалу оценивания.

Вероятность же случайного ввода правильного ответа в этом случае рассчитывается в зависимости от вариантов представления списка элементов последовательности.

Согласно *первому варианту*, все N элементов последовательности входят в правильный ответ. В этом случае вероятность случайного ввода правильного ответа вычисляется из уравнения

$$q_6 = \frac{1}{N!}. \quad (13)$$

Второй вариант рассматривает ситуацию, когда из N предложенных элементов только k входят в правильную последовательность:

$$q_7 = \frac{1}{\sum_{k=2}^N N!}. \quad (14)$$

Задания в открытой форме принципиально отличаются от заданий в закрытой форме. При оценке заданий открытого типа в самом простом случае может быть использована дихотомическая шкала. Однако особую сложность в данном случае составляет проблема проверки правильности свободно-конструируемых ответов. Одним из способов, решающих эту проблему, является так называемый *дескрипторный контроль* – контроль по ключевым словам. В этом случае описание множества правильных ответов включает в себя перечень дескрипторов, связей между ними и схемы (формулы) допустимых ответов. Здесь под дескриптором будем понимать слово или устойчивое словосочетание, служащее для обозначения класса условно эквивалентных ключевых слов и словосочетаний [5]. Остальные близкие по смыслу ключевые слова являются синонимами. Между ключевыми словами и дескрипторами могут также задаваться отношения: род – вид, часть – целое. При перечислении всех возможных последовательностей ключевых слов и описании схем допустимых ответов следует использовать минимальное количество дескрипторов. В более сложных или, иными словами, более «интеллектуальных» программах компьютерного тестирования при анализе ответов строятся концептуальные сети вопроса и сообщения тестируемого. Затем происходит сравнение сети вопроса и сети ответа. И если сеть ответа совпадает с сетью вопроса или, являясь частью сети вопроса, включает в себя необходимые дескрипторы, то ответ признается верным. При сравнении сетей эталонного ответа и ответа студента могут встретиться различные ситуации:

1) сети абсолютно различные, т. е. пересечения сетей нет;

2) сети имеют общие фрагменты (в данном случае применение дихотомической системы оценивания будет неоправданным), что возможно в двух случаях:

– множества дескрипторов сетей пересекаются, а связей нет;

– пересекаются множества и дескрипторов, и связей.

При оценивании вероятности ввода правильного ответа во время выполнения многошагового тестового задания первоначально необходимо оценить вероятность угадывания правильного ответа q_{s_i} на каждом из l шагов.

Заключение. Таким образом, в статье рассмотрены различные методы оценивания знаний, как по отдельному тестовому заданию с учетом его индивидуальных параметров, так и по всему компьютерному тесту в целом. Все из исследуемых методов достаточно просты с

точки зрения программной реализации и эффективны. Однако не менее важная роль принадлежит ряду других факторов, требующих отдельного рассмотрения и исследования. Результаты выполненного анализа планируется использовать при реализации тестовой части инструментального средства автоматизированного проектирования компьютерных обучающих программных модулей «Сфера».

Литература

1. Белоус, Н. В. Моделирование процесса проведения и оценивания практикумов по компьютерной дискретной математике с использованием адаптивного тестирования / Н. В. Белоус, И. В. Куцевич, Н. Н. Куцевич // Матем. системы и машины. – 2009. – № 3. – С. 177–187.

2. Карпова, И. П. Исследование и разработка подсистемы контроля знаний в распределен-

ных автоматизированных обучающих системах: дис. ... канд. техн. наук: 05.13.13 / И. П. Карпова. – М., 2002. – 200 л.

3. Белоус, Н. В. Дифференциальное оценивание знаний при дистанционном тестировании / Н. В. Белоус, И. В. Куцевич // Искусственный интеллект. – 2009. – № 1. – С. 63–73.

4. Дидактические тесты: технология проектирования: метод. пособие для разработчиков тестов / Е. В. Кравец [и др.]; под общ. науч. ред. А. М. Радькова. – Минск: РИВШ, 2004. – 128 с.

5. Савельев, А. Я. Подготовка информации для автоматизированных обучающих систем: метод. пособие для преподавателей и студентов вузов / А. Я. Савельев, В. А. Новиков, Ю. И. Лобанов. – М.: Высш. шк., 1986. – 176 с.

Поступила в редакцию 31.03.2010