

# СТРУКТУРА ЭКСПЕРТНОЙ СИСТЕМЫ ПРОВЕРКИ ЗНАНИЙ ПО РЕЗУЛЬТАТАМ ТЕСТИРОВАНИЯ

*Казаченко Е.В., Фоотов А.М.*

*Донецкий национальный технический университет г. Донецк*

*Кафедра автоматизированных систем управления*

*E-mail: [kazkate@gmail.com](mailto:kazkate@gmail.com)*

## *Аннотация*

*Казаченко Е.В., Фоотов А.М. Структура экспертной системы проверки знаний по результатам тестирования. В статье рассмотрены методы оценки знаний, их достоинства и недостатки, а также возможность применения нечеткой логики к оцениванию знаний по результатам тестирования.*

**Общая постановка проблемы.** В настоящее время для оценки знаний чаще всего применяется стандартная схема индивидуального оценивания методом экспертных оценок, в которой оценка знаний проводится экспертом либо группой экспертов.

Этот подход имеет серьезные недостатки: субъективизм, заключающийся в том, что разные преподаватели могут по-разному оценить способности одного и того же студента; отсутствие широкой шкалы оценивания; «локальность» оценки, которая имеет смысл только в рамках небольшой группы оцениваемых; трудоемкость массового тестирования; задание, как правило, не охватывает весь предмет, что не позволяет оценить реальные знания испытуемого.

Актуальность проблемы заключается в том, что традиционные методы контроля не позволяют достаточно объективно определять качество и уровень знаний, что требуется для внешнего тестирования абитуриентов или тестирования при приеме на работу.

Тестовые задания по своей форме могут быть нескольких типов: на выбор; на соответствие; на ранжирование; на конструирование; ситуационные.

Несмотря на такое многообразие форм тестовых заданий в автоматизированных системах тестирования чаще всего используются простые алгоритмы формирования итоговой оценки: аддитивные алгоритмы и аддитивные алгоритмы со штрафными баллами [3].

Для контроля оценки результатов тестирования применяются адаптивные и неадаптивные методы контроля знаний [1]. При неадаптивных методах в процессе контроля все студенты проходят одну и ту же, заранее определенную, последовательность кадров проверочных заданий, которая не зависит от действий обучаемого во время контроля.

Адаптивные методы максимально используют данные из модели студента (уровень подготовленности студента, уровень беспокойства-тревоги, правильность ответа и др.) и/или модели учебного материала (взаимосвязи между проверяемыми понятиями). Применение адаптивных методов формирования тестов дает возможность более точного оценивания [3].

Разработанная на Западе теория создания тестов (IRT [3, 5]) предназначена для оценки латентных параметров испытуемых и заданий теста. Основным принципом IRT является установление вероятностей связи между наблюдаемыми результатами тестирования и латентными параметрами испытуемого и заданий теста [4]. Эта связь выражается в виде:

$$P(x_{ij}=1) = f(\theta_i - \beta_j) \quad (1)$$

где  $x_{ij}$  - элемент матрицы ответов, равный 1, если ответ  $i$ -го испытуемого на  $j$ -е задание верный, 0 - в противном случае;

$\theta_i$  - уровень подготовки  $i$ -го испытуемого,  $i=1..N$ ;

$\beta_j$  - трудность  $j$ -го задания,  $j=1..n$ ;

$f$  - логистическая функция, зависящая от выбранной модели IRT.

Среди моделей IRT различают однопараметрическую модель Раша, двухпараметрическую модель Бирнбаума, трехпараметрическую модель Бирнбаума.

Оценка знаний может проводиться с использованием различных моделей оценивания: с учетом только правильности ответов студентов или учитывающих параметры заданий и уровень усвоения знаний. Для реализации моделей оценки знаний применяются линейные алгоритмы, экспертные системы, нейронные сети, аппарат нечеткой логики [4].

Несмотря на достаточно большое количество работ по рассматриваемой теме [1, 4, 5], можно выделить ряд общих недостатков современных методов автоматизированного оценивания знаний по результатам тестирования:

- использование одной методики составления теста, что сужает возможности тестирования;
- негибкость процедур расчета итоговой оценки, вследствие применения методов использующих алгоритм накопления баллов, методов ранжирования, методов поощрения и штрафов;
- высокая трудоемкость формирования высокоэффективных тестов, или сведение процедуры формирования теста к случайному выбору вопросов;
- использование заранее сформированных тестов, что исключает возможность использования адаптивных методик тестирования или делает их недостаточно гибкими.

**Постановка задач исследования.** Целью работы являются снижение трудоемкости составления адаптивных тестов за счет внедрения системы автоматической генерации тестов; повышение качества оценивания; снижение трудоемкости за счет автоматизации процесса проверки результатов тестирования и выставления оценки.

Назначение системы – адекватное оценивание специальных знаний лица проходящего тестирования на основе адаптивного комплексного тестирования по заданной предметной области.

Задачи: разработать экспертную систему оценивания знаний, которая будет позволять автоматически генерировать тесты; проводить тестирование; выставлять адекватную оценку высокой степени точности; формировать пояснения выставленной оценки и рекомендации по углублению знаний в той или иной области; давать рекомендации разработчикам тестов по качеству тестовых заданий и тесту в целом [2].

**Решение задачи и результаты исследований.** Система проверки знаний по результатам тестирования реализуется с применением аппарата нечеткой логики для оценки знаний тестируемого.

Рассмотрим структуру экспертной системы оценки знаний (рис. 1):

- Подсистемы тестирования – реализует интерфейс взаимодействия испытуемого и экспертной системы оценки знаний.
- Блок УПП дисциплин – подсистема, отвечающая за наполнение и хранение информации о дисциплинах, взаимосвязях между темами дисциплин и определяющая степень важности той или иной темы.
- Подсистема оценки качества теста на основе генетического алгоритма – реализует автоматическую генерацию тестов определенного качества.
- Подсистема формирования адаптивных тестов – позволяет задать требуемый уровень качества создаваемых тестов и определить значения параметров создаваемого теста. В качестве таких параметров используется число пропусков в разрезе тем, число выполненных лабораторных работ и качество их выполнения.
- База вопросов – хранилище вопросов по заданной тематике. Через интерфейс Подсистемы формирования адаптивных тестов, преподаватель получает возможность формировать тестовые вопросы различных типов, задавать взаимосвязи между ними, задавать положительный либо отрицательный вес каждого ответа на сформулированный вопрос.



Лингвистическая переменная задается терм-множеством  $T$ , универсальным множеством  $X$ , множеством синтаксических модификаторов  $G$ , функцией принадлежности  $F$ .

Множество термов для лингвистической переменной «Уровень практических знаний» приведен в таблице 2.

Таблица 2 – Термы для лингвистической переменной «Уровень практических знаний»

Уровень	Шкала	Критерий
Низкий	0 .. 3	При сдаче лабораторных работ студент выполнил необходимый минимум требований.
Средний	3 .. 6	При сдаче лабораторных работ студент выполнил требуемый минимум, хорошо отвечал на поставленные вопросы.
Высокий	6 .. 9	При подготовке к лабораторным работам студент проводил анализ поставленной проблемы. Ответы на дополнительные вопросы были лаконичные и точные.
Повышенный	9 .. 12	При подготовке к лабораторным работам студент проводил анализ поставленной проблемы. Решение поставленных задач было не только верным, но и оригинальным.

Множество термов для лингвистической переменной «Посещаемость» приведено в таблице 3.

Таблица 3 – Термы для лингвистической переменной «Посещаемость»

Уровень	Шкала	Критерий
Плохая	0 .. 4	Пропущено более 30% занятий
Средняя	4 .. 8	Пропуски занятий составили от 10 до 30%.
Хорошая	8 .. 12	Пропущено менее 10% занятий

Множество термов для лингвистической переменной «Знание темы» приведено в таблице 4.

Таблица 4 – Термы для лингвистической переменной «Знание темы»

Уровень	Шкала	Критерий
Плохая	1 .. 2	Значение лингвистических переменных этого типа формируется на первом уровне экспертной системы.
Средняя	2 .. 3	
Хорошая	3 .. 4	
Отличная	4 .. 5	

Для формирования нечетких переменных на основании лингвистических переменных используется треугольная функция принадлежности, значение которой в точке  $X$  вычисляется по формуле (2).

$$MF(x) = \begin{cases} 1 - \frac{b-x}{b-a}, & a \leq x \leq b \\ 1 - \frac{x-c}{c-b}, & b \leq x \leq c \\ 0, & \text{в остальных случаях} \end{cases}, \quad (2)$$

где  $a$ ,  $b$ , и  $c$  - левая граница, точка максимума и правая граница функции принадлежности соответственно.

Для формирования итоговой оценки, которая является выходной лингвистической переменной, применяется база знаний, которая представлена в виде набора продукций и формируется преподавателем (экспертом) для каждой учебной дисциплины. С помощью механизма продукций, который составляет часть системы нечеткого вывода, преподаватель полу-

чает возможность задать зависимости между изученными темами, дифференциально подойти к оценке посещаемости, задать уровень значимости каждой темы курса.

Для итогового формирования оценки используется алгоритм нечеткого вывода Мамдани.

Как уже упоминалось выше, система позволяет генерировать тестовые задания разной сложности. Для автоматического формирования тестовых заданий используется генетический алгоритм, представленный на рисунке 2.



Рис. 2. Простой генетический алгоритм

В качестве фитнес-функции генетического алгоритма была выбрана трехпараметрическая модель Бирнбаума, т.к. она кроме всего вышеперечисленного, также учитывает вероятность угадывания ответа, исходя из формулировки задания. Такая ситуация может возникнуть, например, при неграмотном подборе дистракторов (вариантов ответа) на задание закрытого типа. Зависимость подчиняется следующей формуле:

$$P_{ij} = c_j + (1 - c_j) \frac{1}{1 + \exp(-d_j(\theta_i - \delta_i))}, \quad (3)$$

где  $c_j$  – вероятность угадывания.

Независимо от выбора модели, изучаются также принципы взаимодействия различных параметров друг с другом.

В трехпараметрической модели Бирнбаума вероятности правильного (соответственно, неправильного(5) решения тестового задания равны:

$$p(\vartheta, \delta, y) = \frac{e^{y(\vartheta-\delta)}}{e^{y(\vartheta-\delta)} + 1} \quad (4)$$

$$q(\vartheta, \delta, y) = \frac{1}{e^{y(\vartheta-\delta)} + 1} \quad (5)$$

Пусть тест содержит  $n$  заданий. Будем считать известными не только трудности заданий  $\delta_1, \delta_2, \dots, \delta_n$ , но и дифференцирующие особенности всех заданий  $\gamma_1, \gamma_2, \dots, \gamma_n$ . Сохраним прежнее обозначение для характеристической функции. Тогда логарифмическая функция правдоподобия равна:

$$\ln(F(\vartheta)) = \sum_{i=1}^n x_i y_i (\vartheta - \delta_i) - \sum_{i=1}^n \ln(e^{y_i(\vartheta-\delta_i)} + 1) \quad (6)$$

Необходимое условие максимума функции (6) приводит к уравнению:

$$\sum_{i=1}^n y_i p(\vartheta, \delta_i, y_i) = \sum_{i=1}^n x_i y_i, \quad (7)$$

из которого должен определяться уровень подготовленности испытуемого.

С помощью модели Бирнбаума оценивается качество и сложность теста.

**Выводы.** Разрабатываемая система призвана повысить качество оценивания знаний, формализовать и автоматизировать методику формирования адаптивных тестов со связанными тестовыми заданиями, оценить качество формируемых системой тестов, предоставить возможность оценивания знаний как в общем по дисциплине, так и по отдельным темам в частности.

В ходе анализа существующих методов решения поставленной задачи принято решение использовать адаптивные методы тестирования, как наиболее точные и нацеленные на всестороннее оценивание знаний тестируемого в совокупности с экспертной системой, основанной на теории нечеткой логики.

Для автоматического формирования адаптивных тестов предлагается использовать аппарат генетических алгоритмов.

Экспертная система оценки знаний по результатам тестирования позволит снизить трудоемкость проведения текущих и окончательных контролей знаний. Позволит проводить многопараметрический анализ успеваемости студентов по различным темам учебного курса. Даст возможность выявить темы, усвоение которых вызвало наибольшие трудности.

## Литература

1. Модели и методы адаптивного контроля знаний [Электронный ресурс] / Л. В.Зайцева, Н. О. Прокофьева // Educational Technology & Society – 2004. – № 7 (4). Режим доступа к журн.: – URL: [http://ifets.ieee.org/russian/depository/v7\\_i4/html/1.html](http://ifets.ieee.org/russian/depository/v7_i4/html/1.html).

2. Минин М.Г. Диагностика качества знаний и компьютерные технологии обучения. Томск: Изд-во ТГПУ, 2000. 216 с.

3. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.: ил.

4. Rash G. Probabilistic Models for Some Intelligence and Attainment Tests, 1960, Copenhagen, Denmark: Danish Institute for Educational Research.

5. Экспертные системы: структура и классификация. – URL: <http://www.ssti.ru>