

ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ ДЛЯ СОВЕРШЕНСТВОВАНИЯ АДАПТИВЫХ ТЕСТОВ ПО ИНФОРМАТИКЕ

*Евгений Игоревич Горюшкин,
аспирант Курского государственного университета*

В СТАТЬЕ РАССМАТРИВАЕТСЯ ОДИН ИЗ ВАРИАНТОВ УСОВЕРШЕНСТВОВАНИЯ СИСТЕМЫ КОНТРОЛЯ УЧЕБНЫХ ДОСТИЖЕНИЙ ПО ИНФОРМАТИКЕ ПУТЁМ ПРИМЕНЕНИЯ АДАПТИВНЫХ ТЕСТОВ ВМЕСТЕ С НЕЙРОСЕТЯМИ.

• Нейросетевые технологии • Педагогический тест • Информационная функция • Математическая теория педагогических измерений • Теория адаптированного тестирования •

В настоящее время информатика является неотъемлемой составляющей образования. Это обусловлено многими причинами, в числе которых глобальная информатизация общества, невозможность осуществлять профессиональную деятельность без эффективного использования преимуществ информационных и телекоммуникационных технологий, необходимость критического осмысления все возрастающих объемов информации, с которыми приходится иметь дело современному специалисту. Не случайно информатика и подходы к обучению работе на основе информационных технологий оказываются в центре внимания многих педагогических исследований.

Наряду с этим очень важно выработать подходы к повышению эффективности обучения, формированию у выпускников способности применять полученные знания, умения и навыки в жизни. Это, в свою очередь, делает актуальной задачу создания и внедрения технологий и средств измерения уровня эффективности образования вообще и обучения информатике в частности.

Анализ методов педагогических измерений позволяет сделать вывод о том, что одним из объективных и эффективных методов контроля качества знаний учащихся образовательных учреждений является тестовый метод, основанный на использова-

нии педагогических тестовых материалов. Трудно назвать дисциплину, в обучении которой так или иначе не применялась бы тестовая форма контроля знаний.

Педагогический тест представляет собой систему заданий возрастающей трудности специфической формы, позволяющей качественно оценить структуру знаний и эффективно измерить уровень знаний испытуемых по различным дисциплинам. Под «тестом» будем понимать *инструмент, состоящий из квалитетически выверенной системы тестовых заданий, стандартизированной процедуры проведения и заранее спроектированной технологии обработки анализа результатов, предназначенный для измерения качеств и свойств личности, изменение которых возможно в процессе систематического обучения.* Выявлено, что одним из наиболее оправданных и эффективных методов использования информационных технологий в системе образования является компьютерное тестирование. На сегодняшний день такое измерение становится все более распространенным вариантом тестирования, благодаря чему наблюдается устойчивый рост применения этого метода в самых различных областях.

Все задания в тесте могут иметь различную степень сложности. Вместе с тем те-

сты по одной и той же тематике также могут отличаться степенью сложности. Данная возможность позволяет педагогу дифференцировать обучение. Большинство авторов, особенно зарубежных, не делает различий между понятиями «сложность» и «трудность». То же самое можно сказать и о разработчиках тестов. Однако есть исследователи, которые различают эти понятия в своих работах. Например, А.Н. Захаров и А.М. Матюшкин отмечают, что степень трудности учебного задания не совпадает с его сложностью. Степень сложности учебного материала характеризуется реальной насыщенностью учебного задания и формой его изложения, а степень трудности всегда предполагает соотношение усвоения учебного материала с ранее усвоенным и интеллектуальными возможностями учащихся.

Использование тестов способствует развитию внимания, памяти, самоконтролю, рассуждению и анализу, обобщению знаний. Тестовые задания должны составляться с максимальным использованием проблемного подхода. Как правило, в тесте должны быть собраны такие задания, которые обладают системообразующими свойствами (т.е. принадлежат к одной учебной дисциплине), связаны и упорядочены по мере возрастания их сложности. Среди основных преимуществ заданий в тестовой форме по отношению к задачам и вопросам можно выделить краткость, логическую структуру, стандартизованность и единую простую процедуру проведения измерений с последующей оценкой их результатов. Тесты могут применяться не только для текущего оценивания знаний тестируемых, но и для итогового контроля знаний.

Яркий пример современного использования итогового контроля знаний учащихся – Единый государственный экзамен (ЕГЭ). На протяжении последних лет Министерство образования и науки РФ активно проводит политику по его внедрению в образовательный процесс. До сих пор учёные-педагоги не сошлись во мнении по поводу использования ЕГЭ для итогового оценивания результатов обучения школьников. Более того, ряд учёных выступает с требованиями отмены проведения ЕГЭ. Несмотря на то, что ЕГЭ применяется для оценивания уровня знаний у школьников, а не у

студентов, его основу составляют те же контрольно-измерительные материалы (КИМ), из которых состоят и тесты. Соответственно, к ним применяются такие же требования и подходы к составлению, как и к заданиям, разработанным для тестов. Это означает, что «КИМы», разработанные для ЕГЭ, могут быть использованы и для проверки уровня знаний студентов. Таким образом, это позволяет отнести ЕГЭ к числу современных примеров итогового оценивания знаний.

В математической теории педагогических измерений (Item Response Theory) существует понятие «информационная функция». График этой функции свидетельствует о дифференцированной точности измерений испытуемых разного уровня подготовленности.

Применение такой информационной функции для оценки качества «КИМов» ЕГЭ выявляет их непригодность для качественного измерения – слабых, средних и хорошо подготовленных испытуемых. В некоторых «КИМах» особенно высока ошибка измерения на крайних уровнях подготовленности: сравнительно эффективно измеряются абитуриенты среднего уровня, но не решается задача качественной оценки слабых и сильных испытуемых. В других «КИМах», где стремились полностью реализовать политические установки Гособнадзора, получилось довольно много ошибок измерений. Слабые выпускники школ вообще не измеряются с необходимой точностью. Именно об этом и говорят данные публикации метрических результатов ЕГЭ¹.

Нынешний облик ЕГЭ ориентирован на некий «средний уровень» учащегося, который заведомо не может соответствовать даже «слабой четверке», это уровень «тройки». Данное свойство нынешнего ЕГЭ объективно, это свойство «большой системы», которая «стабилизирует средний результат». Ориентация ЕГЭ на указанный средний уровень выражается и в структуре ЕГЭ, и в подборе тестовых заданий, и в шкалировании результатов (в выведении оценки из набранных так называемых сырых баллов). Таким об-

¹ <http://www.press.alledu.ru> Вятчанин Н. Контрольно-измерительный выстрел в ЕГЭ.

разом, в ЕГЭ изначально заложены тенденции сдвига «двоечника» к «тройке», а «отличника» к «четвёрке». Конечно, самые «края спектра» получают объективные оценки: безнадёжные двоечники – «двойки», а победители предметных олимпиад – «пятёрки». Но эти крайности имеют отношение скорее к врождённым чертам личности, а не к образованию.

Исследования большого количества разработанных тестов по информатике выявили ряд недостатков, в числе которых не объективность весов тестовых заданий, малое или большое количество тестовых заданий или одновариантность теста, присутствие связи между последовательными заданиями.

На практике часто возникает ситуация, когда одна и та же группа учащихся без особого труда справляется со всеми тестовыми заданиями или, наоборот, не может справиться с большей частью тестовых заданий. Существует проблема корректности подбора сложности тестовых заданий для адекватной оценки уровня знаний учащихся. В связи с увеличением количества и недостаточным качеством тестов, применяемых в обучении информатике в учебном заведении, не всегда представляется возможным качественно определить уровень учебных достижений учащегося, основываясь только на тестах, в которых количество заданий фиксировано (так называемые тесты фиксированной длины).

Устранение этого недостатка возможно за счет применения такого вида тестирования, которое способно «подстраиваться» под уровень знаний студентов, варьировать сложность и количество заданий в зависимости от правильности ответов на них. Таким образом, процесс тестирования адаптируется к уровню знаний испытуемого, что позволяет получить более достоверные результаты, сократить время, затрачиваемое на прохождение теста, не снижает мотивацию учащихся к обучению и тестированию. Подобные подходы принято называть *адаптивным тестированием*, под которым понимается совокупность тестовых заданий, обладающих необходимыми характеристиками для эффективного оценивания уровня и качества подготовленности испытуемого, порядок предъявления кото-

рых зависит от правильности ответа на предыдущие задания.

При компьютерном адаптивном тестировании по информатике тестовые задания формируются индивидуально для каждого экзаменуемого с учетом результатов выполнения предыдущих заданий. Типы заданий, их количество и порядок следования индивидуальны. Таким образом, адаптивное тестирование не только даёт более объективную оценку знаниям, умениям и навыкам обучаемых, но и позволяет выявлять, какие знания ошибочны или неполны, а также формировать дальнейшую траекторию обучения.

Эффективность адаптивных тестов по информатике повышается при использовании многошаговой стратегии отбора и предъявления заданий, основанной на алгоритмах с полной контекстной зависимостью, в которых очередной шаг совершается только после оценки результатов выполнения предыдущего шага.

Сопоставление различных видов адаптивного тестирования по информатике с позиции их эффективности для решения проблем контроля в образовании приводит к выбору в качестве основного подход компьютерного адаптивного тестирования, основанный на многошаговых варьирующих стратегиях. Наиболее важное преимущество варьирующих стратегий связано с возможностью оперативного реагирования на результаты выполнения учебных заданий по информатике путём переоценки уровня подготовленности обучаемого после выполнения каждого очередного задания адаптивного теста. Последнее обстоятельство порождает уникальные возможности в решении проблем индивидуализации при обучении и контроле по информатике.

Благодаря развитию теории адаптивного тестирования стала возможной адаптация не только тестовых заданий, но и тестируемых систем. Эта адаптация может быть направлена на приспособление к предметной области, выбранной для тестирования, а также к текущим потребностям и характеристикам конкретного испытуемого.

Несмотря на все свои преимущества и проведённые педагогические исследова-

ния, адаптивное тестирование имеет ряд недостатков и нерешённых проблем: недостаточно разработаны подходы к оцениванию личностных параметров испытуемых, не полностью исследовано введение аналитических зависимостей между уровнем подготовленности студентов и сложностью заданий при дифференциации режимов адаптивного обучения и адаптивного контроля, имеет место несоответствие запланированной сложности заданий и реальной. В последнем случае, как правило, все задания имеют фиксированные веса (оценка сложности задания). При многократном тестировании складывается ситуация, когда самые сложные задания (по мнению разработчика теста) на самом деле могут иметь сложность не выше среднего.

Для решения этой проблемы необходимо провести переоценивание сложности заданий на основе данных учащимися ответов. Как показывают исследования по интеллектуальному анализу и обработке данных (ИАД), такой тип задач характерен для кибернетического метода ИАД, основанного на технологии нейросетевого анализа данных. Цель ИАД – создание алгоритмического и программного обеспечения для компьютерной техники, позволяющего решать различные интеллектуальные задачи, в число которых можно включить и проблему определения реальной сложности тестовых заданий, необходимых для адаптивного тестирования.

На сегодняшний день искусственные нейронные сети (ИНС) широко применяются в разных областях науки и производства. Обладание такими свойствами, как сбор информации или анализ и принятие решения, если оно требуется, позволяет применять аппарат искусственных нейронных сетей в тестировании для решения проблем, связанных с совершенствованием контрольно-измерительных материалов (КИМ) тестов по информатике, что способствует повышению достоверности результатов тестирования.

Искусственные нейронные сети представляют собой информационные структуры, лежащие в основе функционирования некоторых компьютерных программ. Такие сети состоят из простых однотипных эле-

ментов (нейронов), связанных между собой определённым образом, функциональные возможности которых аналогичны большинству элементарных функций биологического нейрона. Каждый нейрон характеризуется своим текущим состоянием (возбуждён или заторможен). Наглядно ИНС можно представить в виде ориентированного графа, вершины которого будут соответствовать нейронам, а дуги, соединяющие вершины, – синаптическим связям или весам. ИНС обладают рядом ценных свойств, в числе которых способность обучаться, анализировать поступающую информацию, обрабатывать большой объем данных, представленных в разных шкалах, отсеивать ненужную информацию, осуществлять прогноз, параллельно обрабатывать сигналы, благодаря объединению большого числа нейронов в слою. В связи с этим выделяют различные классификации ИНС, модели которых способны решать различные задачи.

Большинство моделей искусственных нейронных сетей при решении любой задачи необходимо обучить, то есть настроить определённым образом весовые коэффициенты. Выделяют три вида обучения: с учителем, без него (самообучение) и смешанное.

В первом случае обучение предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый идеальный выход. Вместе они называются «обучающей парой». Множество таких обучающих пар принято называть «учителем». Обычно сеть обучается на некотором числе таких обучающих пар. Проявляется выходной вектор, вычисляется выход сети и производится сравнение с соответствующим целевым вектором, разность (ошибка) с помощью обратной связи подаётся в сеть, и веса изменяются в соответствии с алгоритмом, стремящимся минимизировать ошибку. Обучение ИНС с учителем подходит для анализа на предмет соответствия задания теста по информатике заявленной тематике.

Во втором случае ИНС не нуждается в целевом векторе для выходов и, следовательно, не требует сравнения с предпо-

делёнными идеальными ответами. Обучающее множество состоит лишь из входных векторов (в нашем случае каждый вектор представляет собой количество верных и неверных ответов на тестовое задание). Обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, т.е. чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. Этот вариант обучения ИНС подходит для анализа уровня сложности тестовых заданий.

При смешанном обучении часть весов определяется посредством обучения с учителем, в то время как остальная часть получается с помощью самообучения.

Таким образом, за счет использования подобного метода интеллектуального анализа данных появляется возможность предварительной разработки тестов по информатике с присвоением заданиям примерных уровней сложности.

Одним из возможных вариантов решить проблему несоответствия заявленной трудности задания реальной является использование перцептрона Розенблатта. От того, насколько удачно выбрана структура ИНС, в значительной степени зависит её способность к обучению, адекватность модели и количество итераций проверки этих свойств, которое предстоит выполнить на пути реализации поставленных задач. Следует помнить, что если сеть не обучается или обучается слишком долго (присутствует эффект переобучения), то это означает, что структура ИНС выбрана неверно. Анализ поставленной проблемы показал, что для её решения достаточно использовать трёхслойный перцептрон.

Одна из первых искусственных сетей, способных к перцепции (восприятию) и формированию реакции на воспринятый стимул, – PERCEPTRON Розенблатта². В данном случае перцептрон рассматривался не как конкретное техническое вычислительное устройство, а как модель работы мозга.

Согласно описанию классического перцептрона Розенблатта и в соответствии с решаемой проблемой одна из возможных моделей нейронной сети, позволяющей модифицировать коэффициенты трудности заданий, если это необходимо, может иметь следующий вид: ИНС состоит из трёх «ветвей», так как будут использоваться три уровня определения трудности задания. Если уровень трудности задания равен единице, то первая «ветвь» будет иметь следующий вид (см. рис 1).

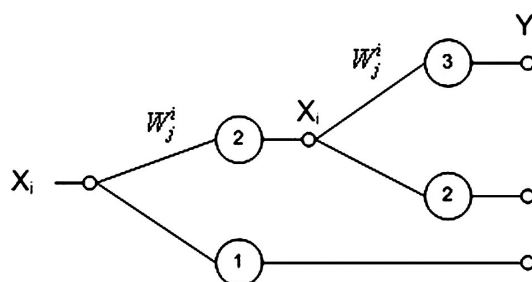


Рис 1. Вид перцептрона при уровне трудности задания, равном 1

Если уровень трудности задания равен двум, то вторая «ветвь» будет иметь следующий вид (см. рис 2).

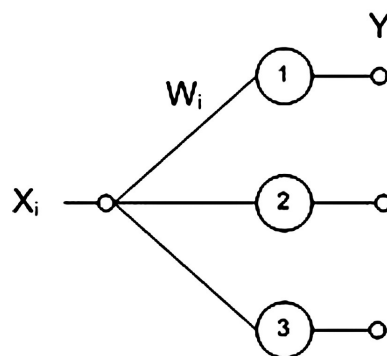


Рис 2. Вид перцептрона при уровне трудности задания, равном 2

При уровне трудности задания равном трём, третья «ветвь» будет выглядеть следующим образом (см. рис 3).

Таким образом, вся модель ИНС (перцептрон) для совершенствования КИМ по информатике будет иметь следующий вид (см. рис 4).

Построенную модель ИНС необходимо обучить на векторе значений (ответы уча-

² Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга. М.: Мир, 1965. 321 с.
Rosenblatt F. Principles of neurodynamics // Spartan., Washington, D.C. 1962. – P. 49.

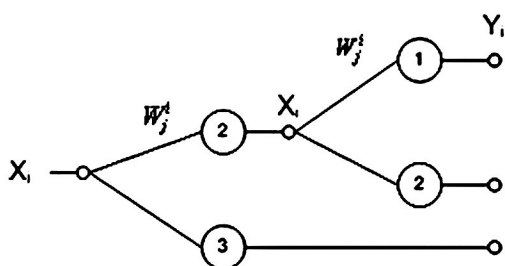


Рис 3. Вид персептрона при уровне трудности задания, равном 3

щихся). Обучение ИНС будет осуществляться методом соревнования, то есть «победитель получает всё». При соревновательном обучении выходные нейроны соревнуются между собой за активизацию. Это явление известно, как правило «победитель берет всё». Подобное обучение имеет место в биологических нейронных сетях. Обучение посредством соревнования позволяет кластеризовать входные данные: подобные примеры группируются сетью в соответствии с корреляциями и представляются одним элементом. При обучении модифицируются только веса «победившего» нейрона. Эффект этого правила достигается за счет изменения сохраненного в сети образца (вектора весов связей победившего нейрона), при котором он становится чуть ближе к входному примеру. Изначально все веса

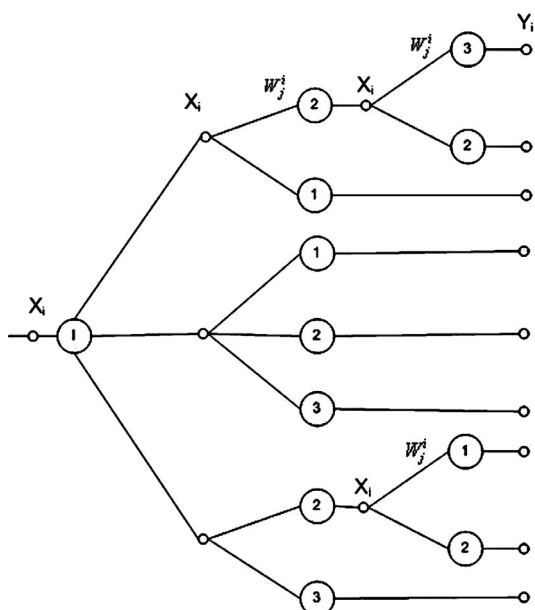


Рис 4. Вид персептрона для совершенствования КИМ по информатике

W_{ij} принимают значение равное нулю. Затем подается вектор значений x_i , состоящий из {0 и 1}. Этот вектор формируется на основе ответов студентов на вопросы из банка заданий, т.е. общее количество нулей или единиц зависит от общего числа ответов, данных на задание. Количество нулей соответствует числу неправильных ответов, а единиц – правильных.

Рассмотрим три варианта обучения персептрона:

Если изначально трудность вопроса равна единице, тогда:

на вход подаётся обучающий вектор, состоящий из единиц и нулей:

при $x_i = 0$, $w_1^{(1)} = w_1^{(1)} + \alpha$; $w_2^{(1)} = w_2^{(1)}$.

при $x_i = 1$, $w_1^{(1)} = w_1^{(1)}$; $w_2^{(1)} = w_2^{(1)} + \alpha$.

если $w_2^{(1)} \geq w_1^{(1)}$, то трудность задания не надо менять, иначе

при $x_i = 0$, $w_1^{(2)} = w_1^{(2)} + \alpha$;

$w_2^{(2)} = w_2^{(2)}$.

при $x_i = 1$, $w_1^{(2)} = w_1^{(2)}$;

$w_2^{(2)} = w_2^{(2)} + \alpha$.

если $w_2^{(2)} \geq w_1^{(2)}$, то трудность задания меняем на три, иначе на два.

Если изначально трудность вопроса равна трём, тогда:

на вход подается обучающий вектор, состоящий из единиц и нулей:

при $x_i = 0$, $w_1^{(1)} = w_1^{(1)} + \alpha$;

$w_2^{(1)} = w_2^{(1)}$.

при $x_i = 1$, $w_1^{(1)} = w_1^{(1)}$; $w_2^{(1)} = w_2^{(1)} + \alpha$.

если $w_1^{(1)} \geq w_2^{(1)}$, то трудность задания не надо менять, иначе

при $x_i = 0$, $w_1^{(2)} = w_1^{(2)} + \alpha$;

$w_2^{(2)} = w_2^{(2)}$.

при $x_i = 1$, $w_1^{(2)} = w_1^{(2)}$;

$$w_2^{(2)} = w_2^{(2)} + \alpha.$$

если $w_2^{(1)} \geq w_1^{(1)}$, то трудность задания меняем на единицу, иначе на два.

Если изначальная трудность вопроса равна двум, тогда:

при $x_i = 0$, $w_1 = w_1 - \alpha$; $w_2 = w_2$;

$$w_3 = w_3 + \alpha.$$

при $x_i = 1$, $w_1 = w_1 + \alpha$; $w_2 = w_2$;

$$w_3 = w_3 - \alpha.$$

Таким образом, обучая ИНС на векторе значений, она будет по завершению обучения выдавать правильные весовые коэффициенты трудности задания на представленную выборку.

Для того чтобы сформировать банк заданий для теста, необходимо определиться с численными математическими величинами, отражающими уровень сложности вопроса, вероятность угадывания, ёмкость ответа на вопрос и т.д.

Рассмотрим следующие характеристики:

Ёмкость ответа на вопрос [I]. Данная характеристика показывает количество информации, содержащееся в ответе на вопрос. Изначально будем считать, что для тестируемого все ответы равнозначны. Тогда для того чтобы оценить величину I, необходимо воспользоваться формулой Хартли для равновероятностных событий:

$$I = \log_2 M, \quad (1)$$

где M – количество возможных ответов на вопрос.

Вероятность правильного ответа на вопрос [P_v]. Эта характеристика отражает

знание группы тестируемых по разделу, которому соответствует вопрос-задание.

$$P_v = \frac{N_v}{N}, \quad (2)$$

где N_v – количество правильных ответов, N – количество всех ответов на вопрос.

Вероятность «слепого» угадывания правильного ответа [P_v]. Этот показатель характеризует событие лишь тогда, когда тестируемый не читает текст задания или не понимает его. В этом случае выбор правильного ответа носит случайный характер (угадывание). У данной характеристики нет единой формулы, поэтому для каждого вида задания она будет рассчитываться отдельно.

Сложность вопроса [Q]. Данная величина учитывает статистику ответов тестируемых и их знание вопроса. Для определения сложности вопроса-задания будем применять формулу энтропии К. Шеннона, отражающая неопределённость состояния объекта,

$$Q = -K \cdot \sum_{j=1}^M P(X_j) \cdot \log_2 P(X_j), \quad (3)$$

где K – положительная константа, которую можно использовать для нормирования и выбора единицы измерения Q; M – количество возможных ответов на вопрос; P(X_j) – вероятность j-го ответа³.

Все выше перечисленные формулы имеют лишь общий вид, а для каждого типа задания они будут различны.

Результаты, полученные в ходе применения выше изложенной технологии, могут быть использованы не только при решении вопросов, связанных с измерением уровня знаний, но и при определении педагогических стратегий, которые учебное заведение должно применять при обучении учащихся информатике. Если в ходе адаптивного тестирования по информатике выявлено общее недопонимание учащимися одной из тем учебного курса информатики, то при последующем обучении педагог должен затратить больше учебного времени и уделить большее внимание изучению этой темы. Дополнительное учебное время и другие ресурсы для этого могут быть высвобождены в результате упрощённого рассмотрения тем, известных учащимся в большей степени. Подобные, «понятные

³ Репьев А.В. Программное обеспечение адаптивной системы дистанционного обучения: Дис... кан. тех. наук. – М., 2006. С. 36.

всем», темы курса информатики также выявляются в процессе тестирования.

Компьютерная обработка результатов тестирования с использованием математических моделей тестовой теории позволяет проводить тестирование на больших выборках и обрабатывать полученные результаты, что помогает получать устойчивые значения параметров тестовых заданий и сравнивать обученность студентов по информатике.

Полученные на основе применения современной компьютерной техники, программного обеспечения и искусственных нейронных сетей результаты тестирования по информатике могут служить рекомендацией педагогам, ответственным за корректировку содержания контрольно-измерительных материалов по информатике. Эти результаты необходимо учитывать при ежегодном корректировании рабочих учебных программ и планов с целью повышения общей эффективности учебного процесса по информатике. □