

МОДЕЛЬ СИСТЕМЫ КОМПЬЮТЕРИЗИРОВАННОГО АДАПТИВНОГО ТЕСТИРОВАНИЯ

Печенежский Н.А., Маслов А.А.
Харьковский национальный университет радиоэлектроники,
Харьков, Украина
E-mail: mpechen@cc.jyu.fi

Abstract. The majority of present day testing systems are static by their nature in a sense of test battery structure, and therefore the provided tests are not personalized to an examinee. In this paper Item Response Theory based model for computerized adaptive testing is considered, and its advantages and limitations are discussed. Then we present an idea of integration of IRT based model with the methodology of expert system design.

Введение. Образование в целом, как и высшее образование – это саморазвивающаяся система, имеющая внутренние механизмы непрерывного развития. Одним из самых важных механизмов такого развития является деятельность по педагогическому контролю качества образования.

Тестовый контроль – один из перспективных методов объективной оценки знаний и способностей учащихся. Однако необходимо признать, что тестирование все еще не играет значительной роли в совершенствовании системы образования, хотя его потенциальные возможности осознаны, а прогресс в вычислительной технике и телекоммуникациях за прошедшие 10 лет дал качественно новые технические возможности для его применения в учебном процессе.

Исследователи выделили целый ряд как организационных, так и технических причин тому, что на сегодняшний день тестирование не оправдывает в полной мере надежды, связываемые с автоматизированным обучением и контролем знаний.

В данной работе мы фокусируемся на проблеме статичности большинства существующих систем компьютеризированного тестирования. В этом контексте под статичностью подразумевается жесткая фиксированная структура тестовых вопросов или заданий в батарее, предлагаемой экзаменуемому. Главный недостаток таких систем заключается в том, что ни информация о текущем уровне знаний экзаменуемого, ни получаемые в ходе тестирования ответы не учитываются для определения следующего «наилучшего» вопроса, который бы был оптимальным по трудности и наиболее информативен, а весь тест – наиболее дискриминативным. Ниже, приведены определения информативности конкретного вопроса, дискриминативности теста, а также другие концепции, связанные с вопросами адаптивного тестирования.

Основы адаптивного тестирования. Адаптивное тестирование - разновидность тестирования, при котором порядок предъявления заданий (или, как правило, трудность заданий) зависит от ответов испытуемого на предыдущие задания.

Дискриминативность - дифференцирующая, различающая способность теста в целом или отдельного тестового задания, указывающая на их способность разделять отдельных испытуемых по уровню выполнения. Если все испытуемые дают на тестовое задание один и тот же ответ, то это означает, что данное задание не обладает дискриминативностью.

Функциональным синонимом дискриминативности является информативность - разнообразие ответов испытуемых на данное тестовое задание: если почти все испытуемые дают одинаковый ответ, пункт считается малоинформативным, то есть недиагностичным, не различающим испытуемых между собой.

Системы адаптивного тестирования, базирующиеся на статистических моделях, используют ранжирование заданий по ряду критериев, основным из которых является трудность тестовых заданий – параметр, под величиной которого обычно понимается монотонно убывающая функция от процента испытуемых, давших правильный ответ на данный вопрос: чем меньше испытуемых справились с заданием, тем выше уровень его трудности. Слишком легкие и слишком трудные пункты автоматически оказываются малоинформативными. Поэтому для большинства тестов оптимальными – соответствующие уровню трудности тестовых заданий, при котором правильный ответ дает примерно половина из той популяции испытуемых, на которой должен быть использован тест – считаются пункты среднего уровня трудности.

Трудность вопроса измеряют экспериментально, этот процесс называется пилотированием трудности заданий - определение процента испытуемых, способных дать правильный ответ на задания, в ходе предварительных (пилотажных) экспериментов.

Для классических систем тестирования характерны два критерия качества - надежность и валидность. Для адаптивного тестирования необходим новый критерий – эффективность, которую можно определить как дифференцированную точность измерения испытуемых различного уровня подготовленности.

Если сравнить понятие эффективность с понятиями надежность и валидность, то самое существенное отличие нового понятия от двух традиционных заключается в переходе от усредненного показателя к дифференцированному.

Надежность относится к тесту, состоящему из зафиксированного числа заданий, предъявляемых всем испытуемым; только тогда можно найти коэффициент надежности

теста, как усредненную меру точности измерения. Аналогично, находится и валидность теста.

Эффективный же тест, напротив, предполагает отход от усреднения и от фиксированного для всех испытуемых числа заданий. Число выбираемых заданий меняется в процессе тестирования, в зависимости от ответа каждого испытуемого. Таким образом, эффективное тестирование - это обязательно индивидуализированное измерение знаний каждого испытуемого с помощью оптимального по трудности и минимального по количеству набора заданий.

Вопрос ответные модели. Item Response Theory (IRT) – статистически обоснованная модель, в которой экзаменуемые могут быть описаны с математической точки зрения набором числовых характеристик, которые можно рассчитывать и прогнозировать опираясь на мат модели, связывающие сложность заданий со способностями экзаменуемого. Наиболее широко распространена IRT-модель, определяемая тремя параметрами.

Согласно данной модели вероятность правильного ответа на заданный вопрос есть функция от действительного уровня способности (знаний) Θ и трех параметров, характеризующий заданный вопрос, а именно:

a_i - дискриминирующий параметр

b_i - параметр, определяющий сложность задания

c_i - параметр, определяющий, на сколько легко угадать правильный ответ, в действительности, не зная его.

Каждое задание имеет свой набор значений данных параметров, которые обычно задаются на основе результатов пилотирования тестовых заданий.

Итак, модель формулирует вероятность правильного ответа экзаменуемого на задание i как функцию зависящую от действительного уровня способности Θ и трех параметров и имеющую следующий вид:

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

Логика компьютерного адаптивного тестирования, базирующегося на IRT-модели, может быть описана как итеративный процесс, состоящий из следующих шагов:

- 1) Все вопросы, которые еще не задавались, анализируются, и определяется наиболее информативный из них, для чего рассчитывается информативность для каждого вопроса по следующей формуле:

$$I_i(\Theta) = \frac{P_i'(\Theta)^2}{P_i(\Theta)(1 - P_i(\Theta))},$$

где

$$P'(\Theta) = (1 - c) \frac{a \exp(-a(\Theta - b))}{(1 + \exp(-a(\Theta - b)))^2}.$$

Экспериментально было доказано, что данная формула может быть упрощена и представлена как:

$$I_i(\Theta) = \frac{2.89a_i^2(1 - c_j)}{(c_i + \exp(1.7a_i(\Theta - b_i)))(1 + \exp(-1.7a_i(\Theta - b_i)))^2}$$

2) «Лучший», т.е. наиболее информативный вопрос для конкретного экзаменуемого, предлагается для ответа.

3) После ответа на вопрос пересчитывается и обновляется значение уровня способности Θ , при этом учитываются все предыдущие ответы экзаменуемого.

Предполагается, что тестирование начинается при инициализированном априорно известном Θ . Далее, рассчитав все I_i , на которые уже был дан ответ, пересчитывается значение уровня способности по формуле:

$$\bar{\Theta}_{s+1} = \bar{\Theta}_s + \frac{\sum S_i(\bar{\Theta}_{s+1})}{\sum I_i(\bar{\Theta}_{s+1})},$$

где

$$S_i(\Theta) = (u_i - P_i) \frac{P_i'}{P_i(1 - P_i)}$$

4) Шаги 1-3 повторяются до момента срабатывания одного из правил останова. Одним из таких правил может служить сравнение значения стандартной ошибки с заранее установленным порогом.

Основным преимуществом компьютеризированных адаптивных тестов, построенных на ИРТ-модели, является значительно меньшее количество времени, необходимое, чтобы определить настоящий уровень экзаменуемого с точностью, сравнимой со значением, получаемым при традиционном тестировании, когда испытуемый должен ответить на все вопросы. За счет этого также уменьшается усталость экзаменуемого, что также является важным человеческим фактором, который непосредственно влияет на результаты тестирования. При этом характерно обеспечение точности уровня знаний для широкого круга экзаменуемых, в то время как традиционные тестирование обычно наиболее точны для экзаменуемых со средними способностями.

К наиболее серьезному недостатку данного подхода следует отнести ограниченность

применения рассмотренной методики для многих областей, особенно принимая во внимание тот факт, что данная модель может иметь дело с анализом правильности-неправильности ответа либо степенью его правильности, однако не способна проанализировать качественные ответы, в том смысле, что психолога часто интересует вариант ответа, а не его «правильность».

Интеграция методологии проектирования экспертных систем и IRT модели.

Экспертная система – это вычислительная система, в которую включены знания специалистов о некоторой проблемной области и которая в пределах этой области способна принимать экспертные решения.

Экспертные оценки - это оценки, которые выносит испытуемому опытный человек (эксперт), обладающий определенными знаниями и о предмете оценивания, и об объекте. - на основании наблюдения, беседы, изучения документов, анализа результатов нестандартизованных тестов.

С целью преодоления вышеуказанного недостатка IRT-модели, а именно, перехода от анализа правильности-неправильности либо уровня правильности предыдущих ответов для принятия решения о наилучшей информативности следующего задания к качественному анализу ответов, оставляя при этом возможность количественного анализа, нами предлагается использование методологии построения экспертных систем продукционного типа и ее интегрирование с IRT-моделью.

Ключевой идеей является предоставление возможности тестологу легко создать адаптивный сценарий тестовых заданий. С этой целью можно использовать интуитивный графический пользовательский интерфейс, позволяющий накладывать набор правил типа «ЕСЛИ-ТО» на исходную IRT-модель для определения «лучшего» следующего вопроса. При этом задаваемые правила будут иметь абсолютный приоритет над IRT-моделью. Также возможно задание дополнительных правил останова теста.

Следует отметить, что процесс интеграции потенциально возможен как на макро, так и на микро уровнях.

Предлагаемая технология может дать начало новой организации как тестового контроля знаний, так и учебного процесса в целом, на более высоком научном уровне, а в условиях массового образования – предоставить возможность эффективной практической реализации принципа индивидуализации обучения.