

УДК 004.4+004.9

Е.В. Седаков

Донецкий национальный технический университет, г. Донецк
кафедра систем искусственного интеллекта

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОЦЕНКИ ЗНАНИЙ В ЭЛЕКТРОННЫХ ТЕСТОВЫХ СИСТЕМАХ

Аннотация

Седаков Е.В. Информационные технологии оценки знаний в электронных тестовых системах. Рассмотрены методы улучшения качества оценки знаний в электронных системах тестирования. Выполнен сравнительный анализ предлагаемых моделей ответа. Сделаны выводы касательно возможности создания универсальных способов оценивания.

Ключевые слова: *неполный ответ, расстояние между множествами, система тестирования.*

Постановка проблемы. Электронные системы тестирования в той или иной форме всё чаще используются для принятия решений, важных в рамках профессиональной жизни человека. Более того, они используются на государственном уровне, как, например, внешнее независимое тестирование, результат которого обрабатывает компьютер, а само оно почти гарантировано однажды станет электронным. Необходимо максимально увеличить точность оценивания знаний, не перегружая при этом набор заданий уточняющими или проверочными элементами. Для этого важно статистически подтвердить корректность одной из моделей оценивания неполных ответов, как закрытого, так и открытого типа, понять, необходимо ли создание новых моделей.

Анализ литературы. Денисова рассматривает весь учебный материал, как множество областей знания, с которыми эксперт связывает конечные дискретные наборы тестовых заданий [1]. Каждое из них принадлежит к одному из выделенных ей универсальных типов. Выбор ответа: тестовое задание данного типа представлено вопросом и конечным дискретным множеством ответов на него, связанных с вопросом. Соответствие: тестовое задание представляет собой вопрос и множество подвопросов к нему. Множество вариантов ответа дискретно, конечно, связано с подвопросами. К данному типу тестовых заданий также относятся вопросы на упорядочивание списка. В этом случае необходимость расположения каждого элемента списка в нужном месте рассматривается как подвопрос, множество возможных мест в списке представляет собой множество вариантов ответа. Ввод чисел: тестовое задания представлено вопросом и дискретным конечным множеством ответов, указываемых экспертом. Множество ответов связано с вопросом. Ввод текста:

тестовое задание данного типа представлено вопросом и дискретным конечным множеством ответов на него, связанных с вопросом.

Согласно модели Денисовой, большая часть информации о тестовом материале является нечеткими отношениями между множествами. Эксперт должен определить множества подвопросов, для каждой области знаний и каждого вопроса. После чего он определяет коэффициенты сложности для каждого вопроса. Также эксперт обязан сформировать нечеткие отношения, для определения степени правильности ответов на вопросы. После того, как тестируемый даст ответы на вопросы, будет сформировано условное нечеткое подмножество его ответов в множестве всех ответов. Его average-представление и позволит определить правильность ответа. Для подвопросов вопросов придется определять свои average-коэффициенты.

Данный способ моделирования отношений между вопросами, ответами и правильными ответами представляется достаточно логичным, если взглянуть на него со стороны теории нечетких множеств. Однако при его использовании могут возникнуть проблемы, связанные сразу с двумя сторонами системы тестирования – техникой и человеческим фактором. Во-первых, нечеткие множества – не самый удобный математический аппарат для представления данных в вычислительных системах, так как сама их архитектура склонна к наличию ограничений, четкости, строгости логичности. Лишь нечеткие множества определенной, не слишком большой, мощности можно будет представить списками. Во-вторых, для работы системы постоянно необходимо присутствие эксперта, идеально разбирающегося в материале и хорошо знакомого с теорией нечетких множеств.

Аванесов рассматривает способ оценивания, согласно которому во многих случаях достаточно представлять тестируемому вопросы с тремя вариантами ответов, один из которых верный [2]. Но оценивание таких ответов ведется, например в системе баллов, которая включает лишь -1, 0, 1. Ученик может закончить тест с негативным количеством баллов, если будет выбирать ответа, который есть логически противоположными правильным. Или полагают педагогом более ошибочными. Такие задачи отвечают принципу импликации и лучше всего подходят для проверки знаний относительно причинно-следственной связи между явлениями.

Подход является очень интересным, так как, согласно словам самого Аванесова, позволяет найти в знаниях тестируемых критические ошибки в логических связях, которые иначе не были бы обнаружены. В то же время, на данный момент имеется не так уж и много систем тестирования, которые позволяют «штрафовать» тестируемого. По каким причинам этот подход, являясь уникальным, не является широко востребованным предстоит выяснить в статистическом исследовании.

Исследование Аванесова относительно психологических аспектов тестирования показывают, что очень часто важную роль в создании тестовых заданий для объективной оценки знаний играют «дистракторы» и их

особенности. Это такие варианты ответов, которые не являются верными, но могут отвлечь внимание тестируемого. Хорошо подобранные дистракторы должны выбираться неподготовленными тестируемыми приблизительно в $1/K$ случаев, где K – общее количество ответов. То есть не должны быть очевидно неверными. Если тестовая задача имеет ответы, которые никогда не выбирают, то система тестирования, которая его использует, не может предоставлять объективные оценки по понятным причинам – все расчеты относительно вероятностей угадывания ответов тестируемыми сразу становятся неверными. За счет очевидно неверного ответа задание становится значительно более легким. Если плохой дистрактор «вычеркнуть» и провести пересчет баллов, то, очевидно, в рамках большинства моделей оценивания, количество баллов, которое наберет среднестатистический ученик, уменьшится.

Очевидно, что подбор правильных дистракторов требует серьезного вмешательства эксперта в предметной области в составление заданий, однако это минимизирует вероятность угадывания, которая является одним из важных факторов множества моделей.

Важным является вклад Карповой в развитие моделей оценивания тестовых заданий с частичными ответами. Предложенный ей метод оценивания называется Дельта-методом или Д-методом [3] и не требует активного вмешательства эксперта в составление заданий. Основой Д-метода является функция подобия множества, как обратная функция расстояния между множествами ответов и правильных ответов:

$$\delta_{\text{мн}} = \frac{K_A}{L_E + K'}, \quad (1)$$

где L_E – мощность эталонного множества, K_A – количество элементов из ответа, которые входят в эталон, K' – количество элементов, не входящих в эталон. Эта оценка изменяется в границах $[0,1]$ и уменьшается, как при нехватке элементов в ответе, так и при наличии лишних. Если правильность ответа зависит от порядка элементов, то он представляется в виде списка. Для сравнения списков можно использовать процедуру сортировки, которая состоит в попарной перестановке элементов. Максимальное число перестановок K_n для списка длиной n можно определить через n :

$$K_n = \frac{n(n-1)}{2}, \quad (2)$$

А степень подобия списков $\delta_{\text{сп}}$ определить как:

$$\delta_{\text{сп}} = 1 - \frac{K_i}{K_n}, \quad (3)$$

где K_i – количество перестановок (инверсий) в списке-ответе. Процедура сравнения списков разбивается на два этапа, на первом из которых происходит сравнение списков, как множеств, а на втором исключаются лишние элементы, а остальные упорядочиваются. Общая степень подобия списков является

функцией от величин $\delta_{\text{мн}}$ и $\delta_{\text{сп}}$. Эту функцию необходимо задавать в зависимости от важности компонент оценки степени подобия. В общем случае – это среднее арифметическое. Считается, что данные методы можно распространить на ответ типа множество списков, в виде которого, в свою очередь, можно представлять ответы на естественном ограниченном языке, таблицы.

Цель статьи – провести анализ методов оценивания неполных ответов различных типов и выделить те модели ответа, для которых необходимо провести статистическую проверку.

Постановка задачи исследования. Необходимо выбрать корректные модели для оценивания неполных ответов в электронных тестовых системах, которые будут давать результаты, максимально приближенные к педагогической оценке. После чего встроить их алгоритмические реализации в одну из имеющихся электронных тестовых систем и собрать достаточно статистических данных для сравнения эффективности работы моделей и корректности получаемых результатов.

Решение задач и результаты исследований. В данный момент ведется встраивание программных реализаций выбранных моделей в одну из существующих систем тестирования. В случае, если ни одна из моделей не окажется универсальной, планируется использование интеллектуальных технологий, для обучения системы тестирования выбору правильной модели оценивания в зависимости от типа вопроса, предметной области и других факторов.

Выводы. Проведен анализ методов оценивания неполных ответов различных типов. Результаты анализа показывают, что многие модели позволяют оценивать неполные ответы без использования специальных экспертных знаний о предметной области, однако их корректность необходимо статистически доказать.

Список литературы

1. Денисова И. Ю. Математические модели представления знаний эксперта в информационной системе дистанционного обучения/ Интернет-ресурс. – Режим доступа: [www/ URL: http://cyberleninka.ru/article/n/matematicheskie-modeli-predstavleniya-znaniy-eksperta-v-informatsionnoy-sisteme-distantsionnogo-obucheniya](http://cyberleninka.ru/article/n/matematicheskie-modeli-predstavleniya-znaniy-eksperta-v-informatsionnoy-sisteme-distantsionnogo-obucheniya) – Загл. с экрана.
2. Аванесов В.С. Научные проблемы тестового контроля знаний // Монография. – М.: Исследовательский центр проблем качества подготовки специалистов, 1994. – 135 с.
3. Карпова И.П. Анализ ответов обучаемого в автоматизированных обучающих системах. – Информационные технологии, 2001, № 11. – с.49-55.