

И.П. Карпова

Московский Государственный институт электроники и математики

Анализ ответов обучаемого в автоматизированных обучающих системах

В статье рассмотрены различные формы представления вопросов и ответов, используемые при контроле знаний. Предложены некоторые формальные методы оценки ответов в обучающих системах. Описаны возможные подходы к анализу текстовых ответов и ответов в виде графических изображений. Показаны преимущества использования предложенных методов.

В настоящее время в процесс обучения широко внедряются информационные технологии [1], в частности, автоматизированные обучающие системы (АОС). Под АОС будем понимать организованный на базе ЭВМ комплекс программного и учебно-методического обеспечения, предназначенный для поддержки процесса обучения и предоставляющий пользователю-непрограммисту возможность настройки на произвольную предметную область и произвольную методику преподавания.

Одной из задач, возникающих при создании АОС, является контроль знаний обучаемого. Он обеспечивает обратную связь с обучаемым и предназначен в первую очередь для определения уровня его знаний с целью организации адаптивного управления обучением.

Стандартизованный контроль знаний

По сложившейся традиции в АОС используются методы т.н. стандартизованного контроля знаний [2]. Сущность их состоит в том, обучаемому предлагается выборка специальных заданий и по ответам на нее выносятся суждения о его знаниях. Для этого выделяются некоторые признаки, в соответствии с которыми ответ обучаемого относится к категории правильных или неправильных. Чаще всего это происходит путем указания эталона ответа и определения процедуры сравнения.

В рамках исследований по обучающим системам (ОС) разработаны различные способы представления и анализа ответов обучаемых. Но, к сожалению, авторы существующих ОС редко пользуются результатами этих разработок, реализуя обычно самые простые формы ответов. Общее представление об уровне решения задачи контроля знаний можно составить по материалам многочисленных конференций, проходивших в последние годы (в частности, [3, 4]).

Наиболее распространенная форма ответов – выборочная: вопрос сопровождается несколькими готовыми вариантами ответов, из которых нужно выбрать один, реже – нескольких правильных ответов. Вторым по популярности идет числовой ответ, обычно, как результат решения предложенной задачи. Кроме этого, иногда используется текстовый ответ, но без анализа, т.е. правильным считается ответ, полностью совпадающий с эталоном. Также встречаются отдельные попытки предоставить возможность в ответе устанавливать соответствие между элементами двух списков, например:

Вопрос: Какие термины обозначают одно и то же множество данных:

- | | |
|------------------------|------------|
| 1) групповое отношение | 4) набор |
| 2) элемент данных | 5) кортеж |
| 3) запись | 6) атрибут |

Эталон ответа: 1-4 2-6 3-5

Ограниченность этого представления заключается в том, что чаще всего поддерживается только бинарное соответствие и одинаковая длина списков элементов, между которыми надо установить соответствие. При этом получается следующее: если обучаемый из n пар соотносящихся элементов знает $(n-1)$ пару, то оставшиеся элементы он автоматически соотносит друг другу, и оценка результата смещается в положительную сторону. Простейшая методика определения оценки O заключается в использовании формулы

$$O = \frac{L_o}{L_e}$$

где L_o – количество правильно указанных пар, L_e – общее количество пар. При использовании такой методики график оценки O для трех пар ответов ($k=3$) будет выглядеть так, как показано на рис. 1. Оценка ответа смещается вправо, т.е. может быть завышена.

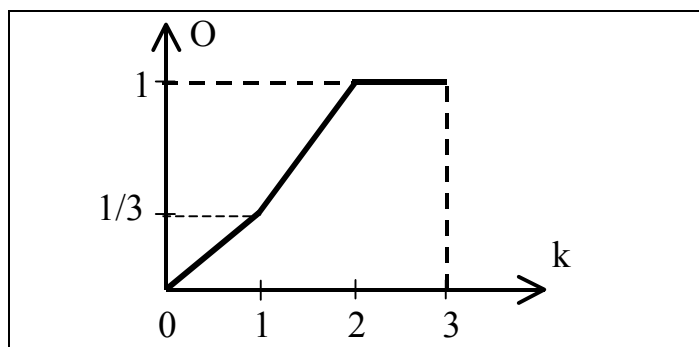


Рис. 1. Оценка ответа на соответствие для трех пар

Примечание: некоторые системы контроля знаний (СКЗ) поддерживают также отдельные специфические формы ответов, например, позволяют создавать структурные схемы на заданной элементной базе или выбирать определенную область на графическом изображении. Но подобные формы ответов не имеют широкого применения, т.к. не обладают достаточным уровнем абстракции.

С точки зрения автора такое положение вещей обусловлено несколькими причинами. Во-первых, сложилось устойчивое представление о том, что все, что выходит за рамки стандартного тестирования, чрезвычайно сложно реализовать. Да и зачем? Стандартное тестирование – метод давно известный, поддающийся статистическому анализу, для него созданы различные математические модели, которые придают этому методу строгость и наукообразность. Во-вторых, усложнение системы контроля знаний обычно ведет к увеличению времени на создание самого теста. А тест при этом остается "одноразовым", т.е. рассчитан на конкретную СКЗ и перенос его в другую оболочку чаще всего является нетривиальной задачей (если вообще возможен).

Но дело в том, что обучающая система должна **адекватно** оценивать **знания** обучаемого (а не здравый смысл, как это часто бывает при выборочном ответе). Кроме того, она должна заставлять человека **думать** над ответом, т.е. выполнять еще и функцию обучения, тем более что при решении реальных задач специалист редко имеет набор готовых вариантов, один из которых – "правильный".

С другой стороны до тех пор, пока не будет разработано математическое и программное обеспечение для общения с ЭВМ на естественном языке без ограничений по предметной области, совсем отказываться от тестирования нецелесообразно. Поэтому тест нужно делать более разнообразным.

Итак, складывается парадоксальная ситуация. Давно разработаны различные формы представления ответов и методы их анализа, но в реальных системах они не используются. В этой статье осуществляется попытка обобщить накопленный в этой

области опыт и представить существующие формы представления ответов и известные (и не очень) методы анализа ответов.

Выбор ответа из множества вариантов

Для начала попробуем отказаться от схемы "один из предложенных ответов – правильный".

Рассмотрим вариант с набором из N готовых ответов. В общем случае правильными среди них может быть произвольное количество ответов: от 0 до N . Если нас не интересует их порядок, то мы получаем *множество* правильных ответов E (эталон). Степень сходства δ_1 ответа обучаемого (A) и эталонного множества E можно определить как

$$\delta_1 = \frac{K_A}{L_E + K'} \quad (1)$$

где L_E – мощность эталонного множества, K_A – количество элементов из ответа обучаемого, входящих в эталонный ответ, K' – количество элементов из ответа обучаемого, не входящих в эталон. Таким образом, при оценке ответа учитывается как наличие "лишних" элементов в ответе, так и отсутствие требуемых, а степень сходства изменяется в пределах $[0,1]$.

Если правильность ответа определяется порядком вариантов, то мы имеем дело со списком. Для сравнения ответа и эталонного списка можно воспользоваться процедурой сортировки [5], которая заключается в приведении одного списка к другому путем попарных перестановок его элементов. Предполагается, что нумерация элементов эталонного списка такова, что он образует полностью упорядоченную подстановку.

Максимальное число перестановок для списка длиной n равно

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

и тогда степень сходства δ_2 списков можно определить как

$$\delta_2 = 1 - \frac{K_i}{K_n} \quad (2)$$

где K_i – количество перестановок (инверсий) в исходном списке (ответе).

Случай с использованием упорядоченного списка вариантов можно еще усложнить, включив в набор готовых вариантов неправильные. При этом процедура сравнения списков будет состоять из двух этапов:

1. сначала списки сравниваются как множества, при этом получается степень сходства δ_1 (1);
2. затем из ответа исключаются все "лишние" элементы (неправильные варианты ответа), а остальные упорядочиваются путем попарных перестановок. Получается степень сходства δ_2 (2).

Получение окончательной величины степени сходства списков δ может быть выполнено, например, одним из следующих способов:

- a) $\delta = (\delta_1 + \delta_2)/2$ – среднее арифметическое δ_1 и δ_2 ;
- b) $\delta = \max(\delta_1, \delta_2)$ – максимум из δ_1 и δ_2 ;
- c) $\delta = \min(\delta_1, \delta_2)$ – минимум из δ_1 и δ_2 ;
- d) $\delta = \delta_1 * \delta_2$ – произведение δ_1 и δ_2 .

Окончательная величина степени сходства остается в прежних пределах $[0,1]$. А графики итоговых оценок для этих четырех вариантов приведены на рис. 2. (ось x – это степень сходства множеств δ_1 , ось y – это степень сходства списков δ_2 , ось z – это общая степень сходства δ).

Даже такая несложная модификация стандартного теста практически сводит на нет вариант "угадывания" правильного ответа.

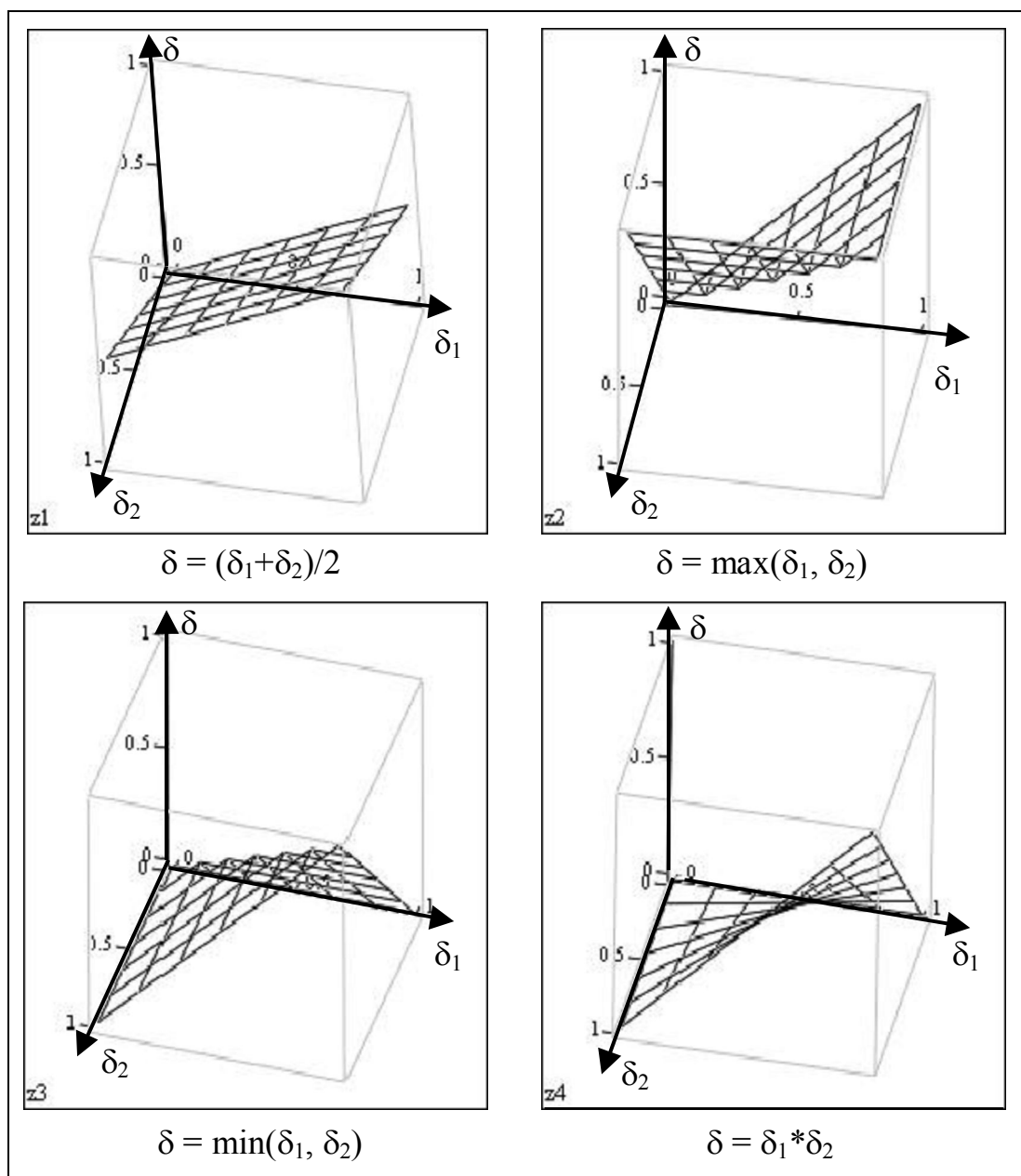


Рис. 2. Графики определения сходства списков δ для разных методов подсчета

Списки множеств и множества списков

Метод анализа списков и множеств можно обобщить на двумерный случай: списки множеств (или списков) и множества списков (или множеств). Это позволит практически на той же алгоритмической базе реализовать ответы типа разбиения на классы и указания соответствия между различными элементами, причем, заметьте, не ограниченные по мощности множеств и длине списков.

Единственное алгоритмическое нововведение при анализе двумерного случая касается множества списков (множеств): для каждого списка нужно сначала выбрать тот элемент из эталонного множества, который будет поставлен ему в соответствие. Это можно сделать, сравнив элементы ответа с каждым из элементов эталонного множества и выбрав "наиболее похожий" (с максимальной степенью сходства). А потом уже сравнивать полученные пары элементов как списки (множества).

В качестве примера использования списка множеств можно привести такой вопрос:

Вопрос: Объедините термины, обозначающие одно и то же понятие, в группу, а группы расположите в иерархическом порядке (от простого к сложному):

- | | | |
|-------------|-----------------|-------------------------|
| 1) запись, | 4) кортеж, | 7) элемент данных, |
| 2) атрибут, | 5) база данных, | 8) агрегат. |
| 3) поле, | 6) набор, | 9) групповое отношение. |

Эталон ответа: {2, 3, 7}, {8}, {1, 4}, {6, 9}, {5}

Предложенный метод, не являясь особенно сложным ни математически, ни алгоритмически, позволяет **формально** оценивать ответы типа вложенных списков и множеств. Более подробно этот метод рассмотрен в статье [6].

Арифметическое выражение

Для ответа типа арифметического выражения должны быть определены эталон (правильное значение) и допустимая погрешность ε . Использование погрешности заключается в том, что если ответ A находится в ε -окрестности эталона E

$$(E - \varepsilon) \leq A \leq (E + \varepsilon),$$

то ответ считается правильным, иначе ответ неверен.

Числовые ответы могут с успехом использоваться в тестировании не только по техническим дисциплинам, но и по гуманитарным. В тех случаях, когда требуется получить точный ответ, погрешность может быть нулевой, например:

Вопрос: В каком году состоялась Куликовская битва ?

Эталон ответа: 1380

Погрешность = 0

Ненулевая погрешность позволяет учитывать ситуации, когда даты исторических событий определены неточно, например:

Вопрос: В каком году родился Томас Торквемада ?

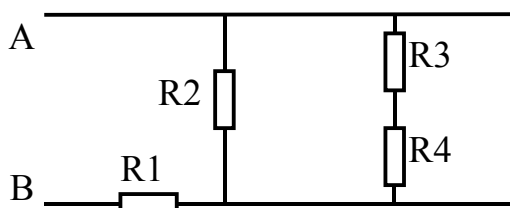
Эталон ответа: 1420

Погрешность = 1

Точная дата рождения великого инквизитора Томаса Торквемады неизвестна. Считается, что он родился около 1420 года. Таким образом, в соответствии с эталоном ответы 1419, 1420, 1421 будут признаны правильными.

Подобная форма записи ответа несравненно лучше, чем выбор из нескольких вариантов или перечисление нескольких эталонов ответа в явном виде. Кстати, перечисление нескольких эталонов ответа встречается не так уж редко, и даже в тех случаях, когда этого можно избежать. Взять, например, ситуацию, когда требуется оценить выражение как формулу. Большое впечатление на автора произвела работа [7], написанная в 1981 году. В ней приведен пример системы контроля знаний, которая позволяет вводить формулы и оценивает их правильность:

Вопрос: Выведите формулу электрического сопротивления между точками A и B.



Эталон ответа: $R1+R2(R3+R4)/(R2+R3+R4)$, $R1+R2(R4+R3)/(R2+R3+R4)$, $R1+(R3+R4)R2/(R2+R3+R4), \dots$

В этой системе (как и во многих аналогичных) требуется задать ВСЕ возможные варианты написания данной формулы, а их для данного ответа как минимум 48.

К сожалению, за последние 20 лет мало что изменилось, если на 2-й Международной выставке-конференции "Информационные технологии и телекоммуникации в образовании", проходившей в Москве на ВВЦ с 6 по 9 апреля 2000 г., автору вновь пришлось столкнуться с системой тестирования по математическому анализу, для которой требовалось задавать в качестве эталона все правильные варианты написания формулы.

И это при том, что еще в 1966 году Ж Питра и независимо от него Дж. Робинсоном был разработан алгоритм унификации [8]. Этот алгоритм позволяет формальным методом определить идентичность любых двух выражений.

Текстовые ответы

Использование ответов, вводимых в свободной тестовой форме, является самой естественной и наиболее сложной задачей при организации системы контроля знаний обучаемых. Анализ текста предполагает использование лингвистических процессоров, выполняющих синтаксический, лексический и семантический анализ текста. Но такие программы на современном уровне развития знаний о языке могут быть ориентированы только на конкретную предметную область. Поэтому, если мы говорим об оболочках АОС, то здесь приходится пользоваться более простыми, но формализованными вариантами.

Наиболее мощной (по сведениям автора) процедурой анализа текстовых ответов из отечественных АОС обладает система AOSMICRO [9]. При задании эталона текстового ответа преподаватель указывает ключевые слова, которые должны присутствовать в ответе обучаемого, и слова, которых там быть не должно. На основании этой информации система анализирует ответ обучаемого и определяет его правильность.

Таким образом, часть работы по анализу текста ответа – разбиение его на словосочетания – выполняется преподавателем. Ответ представляется набором элементов, каждый из которых может быть словом или словосочетанием. Если рассмотреть этот набор как двухуровневую структуру (например, список множеств) и использовать предложенный метод сравнения, то мы получим **формальную** процедуру определения правильности текстового ответа, которая по своим характеристикам не уступает процедуре, реализованной в AOSMICRO.

Другой способ организации работы с текстовыми ответами – шаблоны. Под шаблоном понимается анкета, содержащая произвольное количество полей и пояснения к ним. Приведем пример вопроса с шаблоном ответа:

Вопрос: *Какие виды систем управления базами данных Вы знаете?*

Шаблон: *по степени универсальности* общего назначения и
специализированные

по методам размещения данных централизованные и
распределенные

по модели данных реляционные, сетевые,
иерархические

Пояснения к полям ввода определяют, какая информация должна быть введена в данное поле. Обучаемый может заполнить не все поля, но при этом его оценка, естественно, уменьшится. Для того чтобы избежать ввода слов, неизвестных системе, целесообразно предоставить обучаемому словарь, содержащий все слова, входящие в эталонные ответы для данного теста (или группы тестов).

Графические ответы

Использование графических изображений (рисунков, графиков и т.п.) в качестве ответов значительно расширяет возможности системы контроля знаний обучаемых, а в отдельных случаях является единственно возможной.

Но здесь мы сталкиваемся с классической задачей распознавания образов, которая заключается в том, чтобы классифицировать объект, т.е. определить, что он относится к данному классу и не относится к другому [10]. Решение этой задачи в общем случае разбивается на три этапа: предобработка (кодирование, аппроксимация, фильтрация, восстановление и улучшение объекта), представление объекта (сегментация и выделение производных элементов) и анализ. Содержимое этих этапов зависит от того, какой метод распознавания используется – дискриминантный или структурный.

Эта процедура распознавания является сложной, но можно ее упростить.

Когда речь идет об обучающих системах, распознаваемым объектом является ответ обучаемого на поставленный вопрос. Для того чтобы обучаемый мог дать ответ в графическом виде, необходимо предоставить инструментарий (например, графический редактор). Таким образом, задача сильно упрощается, потому что в качестве элементов графического изображения выступают производные элементы, а операции, производимые над ними, становятся операциями соединения этих производных элементов в изображение (объект). Благодаря этому этап предобработки упраздняется, а этап представления становится чисто техническим. Этап анализа представляет собой простейшую форму распознавания образов – сравнение с эталоном, который также вводится преподавателем через этот графический инструментарий.

Многообразие задач, решение которых может быть представлено в графическом виде, не позволяет рассмотреть все возможные варианты анализа графических изображений или предложить какой-либо универсальный подход к анализу без усложнения как инструментария, так и самих методов анализа изображений. Например, если для качественного анализа графиков достаточно отметить точки экстремумов и характер линий (прямые или дуги), то проверять правильность электронной схемы целесообразнее с помощью функции, которую она должна реализовать.

Качественный анализ графиков

Здесь мы рассмотрим подробнее только задачу анализа графиков как одну из наиболее распространенных.

Когда преподаватель предлагает обучаемому изобразить какой-либо график, чтобы оценить его знания, обучаемый должен правильно отразить на графике точки экстремума и тенденцию графика. Это означает, что анализ графиков должен быть **качественным** и основываться на тех же контрольных элементах: привязка к координатам, точки экстремума, тенденция.

Воспользуемся методом синтаксического распознавания [11]. Сформируем множество производных элементов. Определение производных элементов сильно зависит от характера образов, от предметной области и от набора доступных технических средств. Универсальное решение проблемы выбора производных элементов пока не найдено, но можно предложить несколько общих рекомендаций:

1. Производные элементы должны служить основными элементами образов и обеспечивать сжатое и адекватное описание образов.
2. Выделение и распознавание производных элементов должно быть простым и осуществляться несинтаксическими методами.

Очевидно, что эти два требования противоречат друг другу. Чем сложнее элементы, тем проще описание образа, т.к. требуется меньшее количество структурных связей между элементами. Но сложные элементы нелегко выделить и распознать. Поэтому в реальных задачах приходится искать компромисс между компактностью описания и простотой выделения производных элементов.

В некоторых случаях (и рассматриваемый относится к их числу) требуется, чтобы производные элементы содержали семантическую информацию, важную для конкретного приложения. В нашем примере такой информацией являются координаты. Выделим производные элементы:

- точка экстремума, в которой производная 1-го порядка равна 0;
- линии с разными углами наклона (рис. 3).

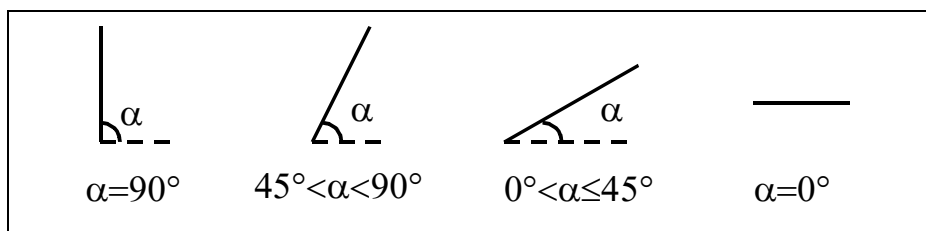


Рис. 3. Производные элементы для задачи качественного анализа графиков

Инструментарий для такой системы должен позволять рисовать прямые линии, позиционировать курсор в точке на линии и отмечать ее координаты. Для формирования эталонного изображения преподавателю предоставляется тот же инструментарий. Оба изображения (эталонное и ответ обучаемого) одинаковым образом переводятся во внутренне представление, где каждому производному элементу ставится в соответствие некоторое мнемоническое и/или цифровое обозначение. В результате объект представляется цепочкой (списком) элементов, а сравнивать списки мы уже умеем. Но при этом возникают некоторые сложности.

График – это двумерное изображение, поэтому описывать его цепочкой символов, используя только операцию конкатенации, неинформативно. Например, если ограничиться указанными производными элементами, то представление образов, приведенных на рис. 4, будет одинаковым.

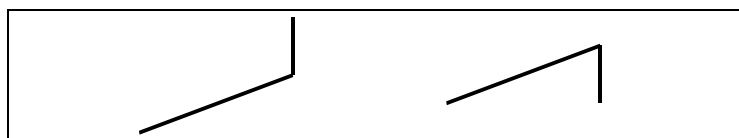


Рис. 4. Примеры графиков

Пути решения возникшей проблемы два: усложнить набор производных элементов или ввести дополнительные операции, указывающие местоположение каждого следующего элемента относительно предыдущего, например, "над предыдущим", "слева от предыдущего" и т.д. Выберем первый путь, чтобы сохранить представление описания в виде цепочки мнемонических обозначений, и дополним набор производных элементов (рис. 5). Таким образом, в производные элементы добавлена синтаксическая информация, а описание образа не усложнилось.

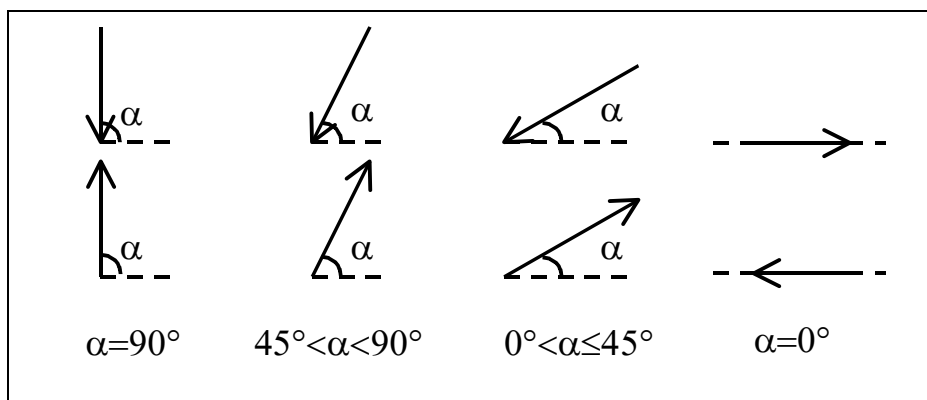


Рис. 5. Дополненное множество производных элементов

Для того чтобы обеспечить возможность анализа графика той же системой, которая обрабатывает списки и множества, можно закодировать линии мнемоническими обозначениями, например, так, как это представлено на рис. 6.

	– DN, вниз ($\alpha=90^\circ$)		– RTUP_BG ($45^\circ < \alpha < 90^\circ$) вправо вверх, большой угол
	– LF, влево ($\alpha=0^\circ$)		– LFDN_BG ($45^\circ < \alpha < 90^\circ$) влево вверх, большой угол
	– LFDN, влево вниз ($\alpha=45^\circ$)		– RTUP_LT ($0^\circ < \alpha < 45^\circ$) вправо вверх, малый угол
	– RTUP, вправо вверх $\alpha=45^\circ$		– RTDN_LT ($45^\circ < \alpha < 90^\circ$) вправо вниз, малый угол

Рис.6. Пример кодирования линий мнемоническими обозначениями

При использовании таких обозначений график функции $y = x^2$ (рис. 7) может быть закодирован следующей последовательностью:

RTDN_BG, RTDN_LT, RTUP_LT, RTUP_BG, которую можно рассматривать как список элементов. (Из списка необходимо удалить повторяющиеся элементы, которые могут возникнуть при попытке изобразить более плавный график).

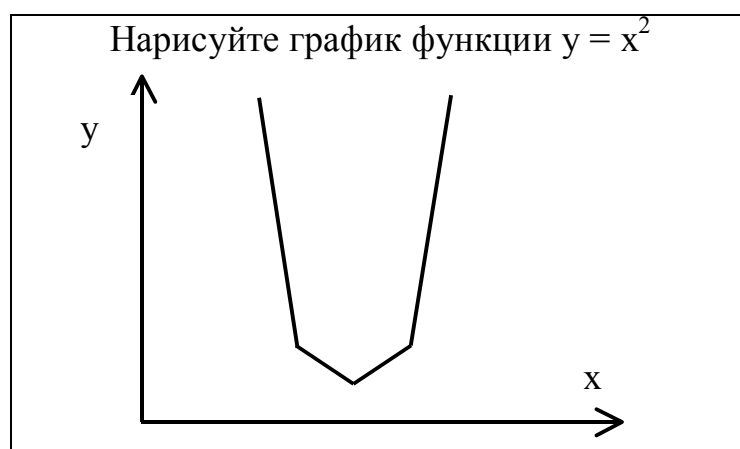


Рис. 7. Изображение графика функции $y = x^2$

Несмотря на чрезвычайную простоту, почти примитивность подобного варианта представления графиков, этот подход дает качественно иной результат по сравнению с выбором одного из нескольких готовых вариантов ответов.

Заключение

Для обеспечения адекватности автоматизированного контроля знаний обучаемых необходимо использовать разнообразные формы представления вопросов и ответов. Предложенные в статье формализованные методы определения сходства множеств и списков позволят:

- повысить надежность тестов, исключив возможность "угадывания" одного правильного ответа из списка предложенных;
- включать в тесты различные по форме вопросы и ответы;
- обеспечить однотипный формальный процесс определения правильности для разных форм представления ответов.

Если же для конкретной предметной области потребуется введение специфических типов вопросов и ответов, то их можно реализовать через внешний процесс, передавая ему информацию о вопросе и эталон ответа и получая взамен оценку ответа обучаемого.

Система контроля знаний, использующая приведенные в статье методы представления и анализа ответов, в настоящее время успешно применяется в Московском Государственном институте электроники и математики для тестирования студентов по различным дисциплинам.

Литература

1. **Шампанер Г., Шайдук А.** Обучающие компьютерные системы // "Высшее образование в России", 1998, №3. – с. 97-99.
2. **Свиридов А.П.** Основы статистической теории обучения и контроля знаний: Метод. пособие. – М.: Высшая школа, 1981. – 262 с.
3. Материалы конференции "Новые информационные технологии в университетском образовании" – Новосибирск: 25-27 марта 1997г. <http://www.nsu.ru>
4. Тезисы докладов уч.-мет. конференции "Современные информационные технологии в учебном процессе" – Ростов: РГУ, 25-26 апреля 2000. <http://www.uic.rsu.ru/~nprohoro/DO>
5. **Кнут Д.** Искусство программирования для ЭВМ / т. 3. "Сортировка и поиск" / Пер. с англ. Под ред. Баяковского и Штаркмана. – М.: "Мир", 1978. – 848 с.
6. **Карпов В.Э., Карпова И.П.** Язык описания системы контроля знаний // "Компьютеры в учебном процессе", 2000, №4. – с. 147-155.
7. **Терещенко Л.Я., Панов В.П., Майоркин С.Г.** Управление обучением с помощью ЭВМ. – Л.: Изд-во ЛГУ, 1981. – 166 с.
8. **Лорьер Ж.-Л.** Системы искусственного интеллекта. / Пер. с франц. – М.: Мир, 1991. – с. 116-161.
9. **Воронин А.Т., Чернышев Ю.А.** Интеллектуальная инструментальная система для WINDOWS // ИТНО'95. Тезисы докладов конференции. Секция В. // <http://petsu.karelia.ru/psu/General/Conferences/Data/19950605>
10. **Фу К.** Структурные методы в распознавании образов. / Пер. с англ. – М.: Мир, 1977. – 320 с.
11. **Ту Дж., Гонсалес Р.** Принципы распознавания образов. / Пер с англ. Под ред. Ю.И. Журавлева. – М.: Мир, 1978. – 414 с.

Karpova I.P. *Pupil answer analysis in automated learning systems*

Some question's and answer's representation forms using in knowledge control are considered. Formal methods of answer evaluation in tutoring systems are proposed. Some methods of analysis for text and graphic answers are described. Advantages of the application of these methods are shown.