

Формирование Стратегии Обучения в Интеллектуальной Обучающей Системе на Основе Обработки Модели Данных и Знаний

А.Р.Ахатов¹, О.Ж.Бобомурадов²

¹Самаркандский государственный университет, Университетский бульвар 15, г. Самарканд,
140104, Узбекистан.

²Институт математики и информационных технологий АН РУз, Дурмон йули 29, г.Ташкент,
100125, Узбекистан

Ozod_b_76@mail.ru, akmalar@rambler.ru

Аннотация. В работе рассматриваются вопросы построения модели состояния обучаемого и определения стратегии обучения, а также формирования полнотекстового учебного материала с обеспечением достоверности текстовой информации на основе методов искусственной избыточности.

Ключевые слова: интеллектуальная обучающая система, обучаемый, стратегия обучения, база данных, база знаний, модель, управление обучением, избыточность, поразрядное суммирование.

1 Введение

Современные информационные технологии находят широкое применение в процессах обучения. Появляются разные взгляды и подходы к их использованию. Развитие информационных технологий способствует быстрым изменениям в сфере обучения, в том числе, за счет интеллектуализации. Построение интеллектуальных систем обучения опирается на новые информационные технологии, включающие в себя разного рода программные средства, методики и нормативы, мультимедийные технологии, визуализацию программных языков, универсализацию компьютерных систем и т.п. Сказанное наглядно может быть показано на примере обучения с использованием электронных учебных пособий [1].

Использование интеллектуальных средств адаптации текущего содержания изучаемого материала (фрагмента электронного учебного пособия) и стратегии управления обучением в соответствии со степенью усвоения изучаемого материала обучаемым является одной из основных задач при разработке компьютерных обучающих систем и электронных учебных пособий.

В связи с этим, интеллектуализация обучающих систем и электронных учебников и пособий, а также организация с их помощью управления процессом обучения является актуальной проблемой.

С целью упрощения реализации содержимое интеллектуального электронного пособия (ИЭП) представляется в системе обучения в виде информационных моделей двух типов: модели данных и модели знаний. В любых информационных системах обработка и достоверность информации является важной компонентой. При формировании полнотекстового материала часто допускаются некорректности ввода информации. В настоящей работе приведены результаты исследований и разработок для описания предметной области системы обучения в модели данных и знаний и обеспечения достоверности информации при представлении материала в базе данных (БД).

2 Принципы построения модели данных и знаний интеллектуальной обучающей системы (ИОС).

Модель знаний – это фреймо-семантическая модель электронного материала, а также правила и процедуры оценки и диагностики уровня усвоения анализа и формирования логических заключений о причинах неправильных представлений обучаемого по соответствующему вопросу, определения стратегий очередного этапа обучения, формирования объёма и содержания учебно-методического материала на очередной этап. Модель знаний хранится в базе знаний (БЗ).

Построение ИЭП в соответствии с рассмотренной моделью предусматривает использование идеологии, методов и средств экспертных систем (ЭС) [2,3].

Модель знаний (МЗ) полного текста ИЭП представляется в виде ориентированного графа

$$\Gamma = I = \{A_i, G_{ik}\}, \quad (1)$$

где A_i – множество вершин, отражающих номера структурно-содержательных частей пособия (частей, разделов, глав, параграфов, подпараграфов и т.д); G_{ik} – множество дуг, характеризующих содержательно-логические взаимосвязи между i -й и k -й вершинами.

В свою очередь, граф полного текста ИЭП состоит из множества подграфов, т.е.

$$I = I_{ji_j},$$

где I_{ji_j} – подграфы, отражающие модели соответствующих содержательных частей ИЭП, т.е. i_j – раздела, j – уровня. Подразделы I_{ji_j} являются частными графами, вложенными в полный граф модели знаний пособия.

В БД ИЭП, как правило, представляется в виде реляционной модели

$$D = \{D_{1t_1}(D_{2t_2}(\dots(D_{jt_j}(\dots(D_{nm})\dots))\dots)\},$$

где D_{ji_j} – полнотекстовая модель i_j – раздела j – го уровня ИЭП (например, i_j – параграф, j – глава).

Между моделями данных (МД) D и знаний в модели полного текста ИЭП (1), а также между их фрагментами обеспечивается взаимно однозначное отображение

$$\begin{aligned} D &\Leftrightarrow I, \\ D_{ji} &\Leftrightarrow I_{ji}. \end{aligned} \quad (2)$$

Для каждого очередного t -го этапа изучения дисциплины ИСО формирует модель I_t требуемого для этого этапа фрагмента ИЭП, логически вытекающего из результатов анализа оценок и диагностики уровня усвоения обучаемым пройденного материала (т.е. оценки ситуации S_t обучения перед t -м этапом), а также выбранной стратегии обучения для t -го этапа (стратегии R_t).

В БЗ формируется множество типовых ситуаций обучения $\{S_t^i\}$, которым поставлены в соответствие множество типовых стратегий обучения $\{R_t^i\}$. Эти отношения отражаются множеством правил продукций

$$\{S_t^i\} \Leftrightarrow \{R_t^i\} \quad (3)$$

Обычно каждой типовой ситуации S_t^i может соответствовать несколько стратегий обучения ($R_t^{i1}, R_t^{i2}, \dots, R_t^{in_i}$). Из них следует выбрать наиболее приемлемую стратегию R_t^* для конкретной текущей ситуации S_t^* , значение которой близко к некоторой типовой ситуации S_t^{ik} . Выбор стратегии R_t^* производится в соответствии с выражением

$$R_t^* = F_R[\alpha(S_t, R_t^i), S_t], \quad (4)$$

где $\alpha(S_t, R_t^i)$ – значение степени предпочтения выбора стратегии R_t^i в ситуации S_t , а сама R_t^* выделяется по наибольшему значению $\alpha(S_t, R_t^i)$ [2,5].

3 Методы определения стратегии обучения в ИОС.

В ИОС представление ИЭП и организация процесса обучения осуществляются на основе суждений, оценок и заключений экспертов, обладающих профессионализмом и мастерством высококвалифицированных педагогов по соответствующим дисциплинам.

На основании заключений эксперта можно определить учебную порцию для обучаемого. Здесь, S_t может выступать как средство, воздействующее на состояние, уровень сложности и состояние среды. Для учета этих средств необходимо определить состояние обучаемого с помощью тестового опроса t_i [2,4].

В БД организуется формирование необходимой порции учебного материала, соответствующего стратегии R_i^* для разных состояний.

Точное значение типа R_i^* для обучаемого находится после выбора соответствующего состояния путем проверки сопоставлений. Если выбранная стратегия R_i^* находится между какими-то (отраженными в базе знаний) R_i^l и R_i^m , то при выборе нужной стратегии формируются множества вида R_i^l и R_i^m .

Расположение этих множеств может быть произвольным, например, заданным в виде множеств $R_i^l = \{x_i\}$ и $R_i^m = \{y_j\}$, где $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ и $y_j = (y_{j1}, y_{j2}, \dots, y_{jn})$ $i = \overline{1, r}$, $j = \overline{1, m}$

Для текущего состояния опорные точки вычисляются следующим образом:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{r}; \quad \bar{y} = \sum_{j=1}^m \frac{y_j}{m}.$$

После определения опорной точки, вычисляется расстояния между элементами множеств:

$$(x, y) = \sqrt{\sum_{i=1}^n (x_i + y_i)^2}$$

После вычисления ω , например в виде $\omega = (\omega, \bar{x}) < (\omega, \bar{y})$, на основе решающего правила уточняем множество:

$$L(x, X) = \min_{i=1, m} L(x, X) \quad (5)$$

Остальные состояния реализуются в заданном порядке.

4 Принципы обработки информации в модели данных.

Модель данных, хранящаяся в БД представляет собой полнотекстовую информацию всего учебно-методического материала по данной дисциплине. Основными задачами модели являются корректное представление максимально полного материала по изучаемой дисциплине и демонстрация материала мультимедиа – средствами. При этом используются технологии, позволяющие извлекать нужную информацию, проводить интеллектуальный анализ данных, создавать средства поиска закономерностей, извлекать знания, анализировать базы данных, т.к. модель данных должна производить обработку полнотекстовых материалов в любых форматах, представленных в виде отсканированных бумажных документов, web-страниц, графических изображений, чертежей, видеофайлов и т.д. Критериями качества работы модели данных являются обеспечение целостности, сохранности и достоверности информационных ресурсов. Соблюдение требований этих критериев, особенно достоверности связано с тем, что при подготовке первичной информации из-за сбоев и отказов в работе технических средств, различного рода помех аппаратных устройств, ошибок операторов информация может стать недостаточно достоверной. Ошибки могут возникать на этапах записи в память ЭВМ, сканирования, перезаписи информации с одного вида носителей информации на другой. Искажения в информации не только снижают точность решения задач, но и влекут за собой ошибочный результат, а следовательно, неверное решение, причем, этап ввода информации считается самой ненадежной частью информационного процесса. Статистикой подтверждается, что наибольшее значение вероятности искажений принадлежит средствам сканирования и распознавания ($\approx 10^{-2}$) и человеку-оператору ($\approx 10^{-3}$) и эти показатели неудовлетворительны для практики обработки информации [1,6]. В связи с этим, разработка методов контроля, направленных на обнаружение и исправление ошибок, в первую очередь, на этапах ввода и подготовки информации, остается до сих пор весьма актуальной и нерешенной задачей.

Для устранения ошибок отмеченных видов проведены специальные исследования для построения теоретической и практической основы применения программных методов контроля достоверности информации [7].

5 Использование искусственной избыточности в методе поразрядного суммирования.

Рассмотрим разработанный нами метод поразрядного суммирования. Будем считать,

что последовательность сообщений состоит из m информационных символов и для контроля достоверности используется n контрольных сумм. Для этого последовательность $a_1, a_2, \dots, a_i, \dots, a_l, \dots$ делится на равные группы, включающие m символов каждая. Количество контролируемых символов определяется по формуле $m = C_n^1 + C_n^2 + \dots + C_n^{n-1}$, где n – число контрольных сумм.

Требуется, чтобы алгоритм контроля информации путем суммирования мог установить не только наличие ошибки в принятой последовательности, но и указать номер позиции, в которой произошла ошибка для автоматического исправления.

Для простоты представления результатов исследований изложим работу метода при $n = 4$. Обозначим через a_i символ i -го разряда в проверяемой последовательности ($i = 1, m$), а через S_j j -ю контрольную сумму ($j = 1, 4$). Будем считать количество контролируемых символов в передаваемой последовательности $m = 14$.

Контрольные суммы передаваемой последовательности данных вычисляются по схеме

$$\left. \begin{aligned} S_1^0 &= a_1 + a_5 + a_6 + a_7 + a_{11} + a_{12} + a_{13} \\ S_2^0 &= a_2 + a_5 + a_8 + a_9 + a_{11} + a_{12} + a_{14} \\ S_3^0 &= a_3 + a_6 + a_8 + a_{10} + a_{11} + a_{13} + a_{14} \\ S_4^0 &= a_4 + a_7 + a_9 + a_{10} + a_{12} + a_{13} + a_{14} \end{aligned} \right\}, \quad (6)$$

в связи с чем образуются новые последовательности символов с учетом контрольных сумм

$$a_1, a_2, \dots, a_i, \dots, a_{14}, S_1^0, S_2^0, S_3^0, S_4^0 \quad (7)$$

По схеме (6) вычисляются контрольные суммы (S_j^0) принятой последовательности и сравниваются с суммами S_j^0 :

$$\Delta S_j = S_j^0 - S_j^0, \quad (j = 1 \div 4).$$

Полученные разности контрольных сумм используются для контроля, обнаружения и исправления ошибочных символов.

Контроль достоверности информации осуществляется на основе пяти правил, изложенных в [8]: π_1 – общее правило контроля достоверности информации; π_2 – обнаружение и исправление однократных транспозиционных ошибок; π_3 – обнаружение и исправление двукратных смежных транспозиционных ошибок; π_4 – обнаружение ошибок типа пропуск или добавление дополнительного символа в строке; π_5 – обнаружение случайных символьных ошибок более высокой кратности.

Исследованы эффективность и предельные возможности метода контроля информации при использовании этих правил. Установлено, что при построении алгоритмов контроля информации по правилам π_2 , π_3 и π_5 ошибки не обнаруживаются в следующих случаях:

- когда одновременно искажаются один, два, три и т.д. символов a_i и соответствующие им контрольные суммы S_j ;

- при искажении значение символа a_i увеличивается на некоторую величину и сумма S_j уменьшится на эту же величину.

В качестве критерия эффективности метода принят критерий вероятности необнаруженных ошибок. Тогда, эффективность алгоритмов, построенных по правилу π_2 равна

$$P_H^1 = 67mnp^2(1-p)^{m+n-2};$$

по правилу π_3 -

$$P_H^2 = 67[C_m^2 C_n^2 p^4 (1-p)^{m+n-4}];$$

по правилу π_5 -

$$P_H^3 = 67[C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}],$$

где k – кратность ошибок ($k > 2$).

По правилу π_4 форматные ошибки не обнаруживаются, когда происходят «пропуск» и «добавление» символов одновременно. Вероятность ошибок типа «пропуск» или «добавление» символов человека-оператора принята равной $P = 2,4 \cdot 10^{-3}$ ош/знак. Учитывая это, получаем вероятность необнаружения форматных ошибок: $P_H^4 = 5,76 \cdot 10^{-6}$ ош/знак.

Суммарная вероятность необнаруженных ошибок рассматриваемых правил контроля информации определяется

$$P_H = P_H^1 + P_H^2 + P_H^3 + P_H^4 = \{67[mnp^2(1-p)^{m+n-2} + C_m^2 C_n^2 p^4 (1-p)^{m+n-4} + C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}\} + 5,76 \cdot 10^{-6}$$

Таким образом, при применении разработанного метода вероятность необнаруженных ошибок составит $P_H \approx 10^{-5} - 10^{-6}$. При этом обнаруживаются и исправляются все ошибки типа однократных транспозиционных и двукратных смежных транспозиционных.

6 Заключение

Предложенная методика реализации функционально – логических и дидактических возможностей ИЭП, по которой полнотекстовый учебный материал, содержащийся в МД, формируется на основе методов автоматического обнаружения и исправления ошибок на этапах ввода и подготовки, а оценка состояния обучаемого и последующее формирование стратегии обучения осуществляются на основе анализа и обработки МЗ, позволяет создать эффективную среду обучения, а также обеспечить достоверную передачу порции учебного материала.

Отметим, что за рамками настоящей работы остается нерассмотренным круг вопросов, связанных с использованием структурно-технологической и программно-временной избыточности для повышения достоверности обработки текстов при формировании МД, а также методов нечеткой кластеризации и распознавания при оценке состояний обучаемого и формировании стратегии обучения.

Литература

- [1] Бобомуродов О.Ж., Ахатов А.Р.: Обработка информации в системах приобретения знаний. Издательство «ФАН» АН РУз, 2009г., 136с.
- [2] Bekmuratov T.F., Bobomurodov O.J., Dadaboeva R.A.: The level of projection and modeling intelligent teaching programs. WCIS-2004, Tashkent, 2004, p.13-16.
- [3] Петрушин В.А.: Интеллектуальные обучающие системы: архитектура и методы реализации. Техническая кибернетика, №2, 1993г.
- [4] Ю.И.Журавлев, М.М.Камилов, Ш.Е.Туляганов.: Алгоритмы вычисления оценок и их применение. Ташкент, изд-во: «Фан», 1974. - 120 с
- [5] Нишанов А.Х., Туляганов Ш.Е., Бобомуродов О.Ж., Худайбердиев М.Х.: Распознавания и адаптация для дистанционного образования. Математические Методы Распознавания образов/Доклады XI Всероссийской конференции. Москва, 2003, с.387-389.
- [6] Ахатов А.Р.: Методы и алгоритмы обеспечения достоверности текстовой информации на основе статистической избыточности. Журнал Вестник ТашГТУ, №2, 2007, ТГТУ, Ташкент, с.41-44.
- [7] Ахатов А.Р., Жуманов И.И.: Алгоритм контроля качества текстов в системах электронного документооборота. Журнал «Вестник ТУИТ», №2/2007, Ташкент, с. 68-72.
- [8] Ахатов А.Р.: Программные методы контроля достоверности в структуре пакетов передачи данных. Естественные и технические науки, Москва, №3, 2008г., стр.310-318.