

СИНТЕЗ ОПТИМАЛЬНОЙ СТРУКТУРЫ ИНФОРМАЦИОННОГО БАЗИСА ДЛЯ ИНТЕРАКТИВНЫХ СИСТЕМ ОБУЧЕНИЯ

Ступина А. А.¹, Корпачева Л. Н.¹, Кузьмич Р. И.², Федорова А. В.¹, Ежеманская С. Н.¹

¹ФГАОУ ВПО «Сибирский федеральный университет», Красноярск, Россия (660041, г. Красноярск, пр. Свободный, 79) e-mail: saa55@rambler.ru

²ФГБОУ ВПО «Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева», Красноярск, Россия (660014, Красноярск, пр. им. газеты «Красноярский рабочий», 31)

Рассматривается задача оптимального разбиения структуры информационного базиса учебно-методических материалов адаптивно-обучающих систем на модули. Показывается, каким образом можно определить оптимальные характеристики информационного базиса при разбиении его на модули. Для моделирования вероятностного процесса обучения отдельного модуля информационного базиса учебно-методических материалов используются концепции цепей Маркова. Исходная информация для модели включает список узлов (шагов процесса обучения), графы связи между ними, матрицу вероятностей перехода от узла к узлу и оценку средней трудоемкости каждого шага. Список шагов процесса, трудоемкости их выполнения в часах и связи между ними определяются структурой информационного базиса. Предложены алгоритмы решения задач определения оптимальной по трудоемкости освоения блочно-модульной структуры информационного базиса для случаев равномерной разбивки базиса и его синтеза из заданных блоков.

Ключевые слова: информационный базис, трудоемкость, цепи Маркова, критерий оптимизации.

SYNTHESIS OF THE INFORMATIONAL BASIS OPTIMAL STRUCTURE FOR INTERACTIVE EDUCATIONAL SYSTEMS

Stupina A. A.¹, Korpacheva L. N.¹, Kuzmich R. I.², Fedorova A. V.¹, Ezhemanskaya S. N.¹

¹Siberian federal university, Krasnoyarsk, Russia (660041, Krasnoyarsk, Svobodnii av. 79), e-mail: saa55@rambler.ru

²Siberian state aerospace university n.a. M. F. Reshetnev, Krasnoyarsk, Russia (660014, Krasnoyarsk, Krasnoyarskii rabochii av. 31)

The optimal subdividing problem into modules of the informational basis structure of educational and methodical materials of adaptive-educational systems is considered. The paper shows the way to determine the optimal characteristics of the informational basis while dividing it into modules. For modeling the probabilistic educational process of the separate module of the informational basis of the educational and methodical materials we use Markov chains concept. The initial information for the model includes a node list (educational process steps), count connection, matrix of the probabilistic transition from a node to node and the average laboriousness estimation of each step. The nodes list of the process, its fulfillment in hours laboriousness, connection are determined by the informational basis structure. The algorithms of the optimal mastering laboriousness of the block-module informational basis structure problems solving for the cases of the uniformly divided basis and its structures from the given blocks are presented.

Keywords: informational basis, Markov chains, optimization criterion.

Введение

Проблема оптимизации формирования информационного базиса (ИБ) в интерактивных адаптивно-обучающих системах, проектируемых на основе современных принципов построения моделей курсов, позволяющих унифицировать требования к их блочно-модульной структуре, последовательности представления компонентов ИБ учебно-методических материалов (УММ), упаковке курсов в уникальные оболочки и среды, является в настоящее время актуальной [3; 4]. В данной работе исследуется зависимость общей трудоемкости прохождения курса от степени дробления его ИБ УММ на блоки и

сборки их в модули УММ. Целью работы является синтез оптимальной по трудоемкости освоения структуры ИБ для интерактивных систем обучения.

Материал и методы

Следуя часто используемой дидактической модели процесса интерактивного освоения ИБ УММ, можно принять гипотезу об экспоненциальном характере зависимости вероятности успешного прохождения контрольных мероприятий от объема базиса. Если объем информационного материала некоторого базиса составляет Θ часов, то вероятность успешного выполнения контрольных мероприятий может быть представлена в виде:

$$p(\Theta) = \exp(-\lambda\Theta), \quad (1)$$

где λ (1/час) – константа, показывающая скорость снижения вероятности успешного завершения процесса освоения материала информационного базиса УММ в зависимости от его объема.

Зависимость (1) соответствует интуитивному представлению об успешности процесса прохождения интерактивного курса. Если объем информационного базиса мал ($\Theta \rightarrow 0$), то вероятность освоения этого базиса стремится к единице, а если он весьма велик ($\Theta \rightarrow \infty$), то вероятность его освоения стремится к нулю. Величина λ зависит от вида (содержательного, структурного и т.п.) базиса, она может быть оценена экспертно или экспериментально на основе опыта работы с определенным видом информационно-терминологических базисов.

Рассмотрим цепь Маркова [5], описывающую вероятностный процесс изучения отдельного модуля информационного базиса УММ (рис. 1).

В данной модели выделены следующие состояния: $S1$ – изучение теоретического материала ИБ УММ; $S2$ – получение справок и консультаций при изучении теоретического материала; $S3$ – выполнение контрольных мероприятий; $S4$ – завершение изучения модуля ИБ УММ.

Зададим вероятности переходов из одного состояния в другое: r – вероятность обращения за справками или консультациями при изучении УММ интерактивного курса; s – вероятность перехода к контрольным мероприятиям; p – вероятность успешного выполнения контрольных мероприятий и завершения освоения модуля курса; q – вероятность неудачи при выполнении контрольных мероприятий и повторного изучения модуля.

В рассматриваемой схеме выполняются очевидные соотношения:

$$r+s=1, \quad p+q=1. \quad (2)$$

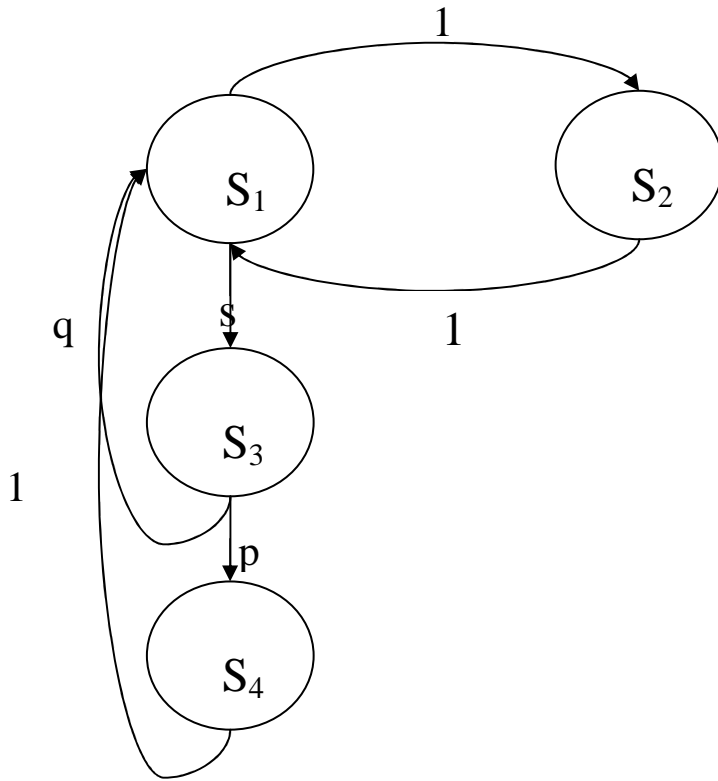


Рис. 1. Модель вероятностного процесса изучения отдельного модуля ИБ УММ

Матрица вероятностей переходов между состояниями:

$$\begin{array}{c}
 \begin{array}{cccc}
 & s1 & s2 & s3 & s4
 \end{array} \\
 \begin{array}{c}
 s1 \\
 s2 \\
 s3 \\
 s4
 \end{array}
 \end{array}
 P = \begin{bmatrix}
 0 & r & s & 0 \\
 1 & 0 & 0 & 0 \\
 q & 0 & 0 & p \\
 1 & 0 & 0 & 0
 \end{bmatrix}.$$

Выделим в P матрицу вероятностей переходов между состояниями невозвратного множества и обозначим Q .

$$\begin{array}{c}
 \begin{array}{ccc}
 & s1 & s2 & s3
 \end{array} \\
 \begin{array}{c}
 s1 \\
 s2 \\
 s3
 \end{array}
 \end{array}
 Q = \begin{bmatrix}
 0 & r & s \\
 1 & 0 & 0 \\
 q & 0 & 0
 \end{bmatrix}.$$

Тогда матрица среднего числа пребывания процесса в состояниях невозвратного множества определяется по формуле:

$$N = (I - Q)^{-1} = \begin{bmatrix} 1 & -r & -s \\ -1 & 1 & 0 \\ -q & 0 & 1 \end{bmatrix},$$

где I – единичная матрица размерности 3×3 .

Поскольку процесс всегда стартует из состояния S_i , нам достаточно определить только первую строчку матрицы N ; ее элементы обозначим n_1, \dots, n_4 . Прделав вычисления обратной матрицы $(I - Q)^{-1}$, получим оценки среднего числа попыток изучения всех компонентов информационного базиса УММ – среднее число пребываний процесса соответственно в состояниях $\{S_1, \dots, S_4\}$ при старте из состояния S_1 :

$$n_1 = \frac{1}{s \cdot p}; \quad n_2 = \frac{r}{s \cdot p}; \quad n_3 = \frac{1}{p}; \quad (3)$$

при этом здесь n_1 – среднее число попыток изучения информационного материала; n_2 – среднее число обращений за справками и консультациями при изучении информационного материала; n_3 – среднее число попыток выполнения контрольных мероприятий.

Обозначим трудоемкости изучения всех компонентов информационного базиса УММ (в часах): Θ – трудоемкость изучения информационного материала; Θ_1 – трудоемкость получения справок и консультаций при изучении информационного материала (состояние S_2); Θ_2 – трудоемкость выполнения контрольных мероприятий (состояние S_3).

Суммарная трудоемкость прохождения информационного базиса УММ с учетом (3) определяется выражением:

$$R = n_1 \cdot \Theta + n_2 \cdot \Theta_1 + n_3 \cdot \Theta_2 = \frac{1}{p} \left[\frac{1}{s} (\Theta + r\Theta_1) + \Theta_2 \right]. \quad (4)$$

Рассмотрим оценки параметров, входящих в формулу (4). Вероятность r определяются экспертно или на основе обработки протоколов прохождения интерактивных учебных курсов. Вероятность p определяется формулой (2). Трудоемкость Θ определяется объемом учебного курса, для которого формируется ИБ УММ, трудоемкость Θ_1 может быть взята как некоторая доля величины Θ :

$$\Theta_1 = k_1 \Theta, \quad (5)$$

где $k_1 < 1$ – доля затрат на получение справок и консультаций, определяемая экспертно или на основе статистики. Трудоемкость Θ_2 может быть представлена как сумма постоянной и переменной составляющих:

$$\Theta_2 = m + k_2 \Theta, \quad (6)$$

где m (часов) – трудоемкость работы по выполнению контрольных мероприятий, не зависит от размеров ИБ УММ; $k_2 < 1$ – доля затрат на выполнение контрольных мероприятий.

В результате подстановки значений параметров (5), (6) формула (4) приобретет вид:

$$R = e^{\lambda \Theta} \left[\left(\frac{1}{s} (1 + k_1 r) + k_2 \right) \Theta + m \right] = R_1 + R_2, \quad (7)$$

причем $R_1 = e^{\lambda \Theta} \left[\frac{1}{s} (1 + k_1 r) + k_2 \right] \Theta$ – затраты на изучение ИБ УММ; $R_2 = e^{\lambda \Theta} (m)$ –

затраты времени на выполнение контрольных мероприятий. Обозначим:

$$\frac{1}{s} (1 + k_1 r) + k_2 = k. \quad (8)$$

С учетом (8) выражение для суммарной трудоемкости (7) упростится

$$R = e^{\lambda \Theta} (k \Theta + m). \quad (9)$$

Рассмотрим задачу разбивки некоторого информационного базиса УММ с общим объемом материала Θ часов на n модулей [1], каждый из которых имеет объем Θ_i часов:

$$\Theta = \sum_{i=1}^n \Theta_i. \quad (10)$$

Трудоемкость изучения каждого модуля в соответствии с (9) составит:

$$R_i = e^{\lambda_i \Theta_i} (k_i \Theta_i + m_i), \quad (11)$$

а общая трудоемкость освоения базиса:

$$R = \sum_{i=1}^n R_i = \sum_{i=1}^n e^{\lambda_i \Theta_i} (k_i \Theta_i + m_i). \quad (12)$$

Рассмотрим предельные случаи. При $n = 1$ получаем формулу (9). При $n \rightarrow \infty$ $\Theta_i \rightarrow 0$, $\exp(\lambda_i \Theta_i) \rightarrow 1$, и из (12) следует, что $R \rightarrow \infty$. Таким образом, функция R имеет, по крайней мере, один минимум.

Выбирая количество модулей n и их объем Θ_i , можно добиться наименьшей общей трудоемкости освоения модульной структуры ИБ УММ. Математически эта задача формулируется следующим образом: задан критерий (12) при условиях (10). Требуется найти такие n и $\{\Theta_1, \dots, \Theta_n\}$, чтобы обеспечить оптимальные значения критерия (12):

$$\hat{R} = \min_{n, \Theta_1, \dots, \Theta_n} R. \quad (13)$$

Задача (13) представляет собой задачу оптимизации нелинейного критерия при ограничениях на переменные. В аналитическом виде эта задача решается достаточно сложно, а численно может быть решена путем перебора вариантов разбивки базиса УММ на модули.

Результаты

Наличие экспоненциальных множителей в выражении (12) говорит о том, что лучших результатов можно ожидать, если величины λ_i и Θ_i будут равны между собой. Рассмотрим задачу о разбивке базиса на модули, равновеликие по объему изучаемого материала.

Положим, что $\Theta_i = \Theta/n$, $i=1, \dots, n$, т.е., что все модули имеют одинаковый объем, кроме того, они все имеют одинаковые параметры $\lambda_i = \lambda$, $k_i = k$, $m_i = m$. Тогда критерий (12) примет вид:

$$R = n \cdot e^{\lambda \Theta/n} (k \Theta/n + m) = e^{\lambda \Theta/n} (k\Theta + mn) \quad (14)$$

и единственным параметром оптимизации становится n .

Для получения оптимального значения n необходимо решить относительно n

уравнение $\frac{dR}{dn} = 0$ или, с учетом (14),

$$\frac{dR}{dn} = -\frac{\lambda \Theta}{n^2} e^{\lambda \Theta/n} (k \cdot \Theta + n \cdot m) + e^{\lambda \Theta/n} \cdot m = 0.$$

Отсюда после некоторых преобразований получаем квадратное уравнение относительно n :

$$n^2 - \lambda \Theta n - \frac{\lambda \Theta^2 k}{m} = 0. \quad (15)$$

Оптимальное значение $n = \hat{n}$ определяется положительным корнем уравнения (15):

$$\hat{n} = \Theta \left(\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + \frac{\lambda k}{m}} \right). \quad \text{Величина} \quad \hat{\Theta} = \frac{\Theta}{\hat{n}} = \left(\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4} + \frac{\lambda k}{m}} \right)^{-1} \quad \text{определяет}$$

оптимальный размер модуля ИБ УММ.

Рассмотрим еще одну практически важную разновидность задачи синтеза структуры информационного базиса для интерактивных систем обучения. Пусть информационно терминологический базис состоит из множества априори заданных блоков $\Sigma = \{S_1, \dots, S_n\}$.

Каждый блок S_i характеризуется параметрами: Θ_i – объем информационного материала блока ИБ, час; m_i – трудоемкость организации контрольных мероприятий, час; λ_i – коэффициент сложности, 1/час; k_i – коэффициент увеличения трудоемкости. Кроме того, для каждой пары блоков $S_i, S_j \in \Sigma$ установлен порядок их изучения: отношение $S_i \rightarrow S_j$ означает, что блок S_i должен быть изучен раньше блока S_j , а отношение $S_i = S_j$ означает, что эти блоки могут изучаться в произвольном порядке.

Задача заключается в следующем: указанные блоки должны быть скомпонованы в модули таким образом, чтобы трудоемкость освоения интерактивного курса в целом была бы минимальной. Каждый модуль может включать один или несколько блоков, причем упорядоченность блоков должна сохраняться как внутри каждого модуля, так и в целом по базису. Требуется сформировать m модулей (m априори неизвестно), каждый из которых содержит множество блоков $\Sigma_j = \{S_{1j}, S_{2j}, \dots, S_{ij}\}, j = 1, \dots, m$, причем

$$\bigcup_{j=1}^m \Sigma_j = \Sigma, \Sigma_j \cap \Sigma_k = \emptyset, \text{ при } j \neq k. \text{ Первое равенство в этом выражении означает, что}$$

должны быть использованы все блоки, а второе – что разные модули не должны содержать одинаковые блоки.

Характеристики каждого модуля определяются на основе характеристик, вошедших в него блоков:

$$\bar{\Theta}_j = \sum_{S_k \in \Sigma_j} \Theta_k; \quad \bar{m}_j = \max_{S_k \in \Sigma_j} m_k; \quad \bar{\lambda}_j = \frac{\sum_{S_k \in \Sigma_j} \lambda_k \Theta_k}{\bar{\Theta}_j}; \quad \bar{k}_j = \frac{\sum_{S_k \in \Sigma_j} k_k \Theta_k}{\bar{\Theta}_j}.$$

Трудоемкость изучения j -го модуля определяется ранее полученной формулой (11):

$$\bar{R}_j = e^{\bar{\lambda}_j \bar{\Theta}_j} (\bar{k}_j \bar{\Theta}_j + \bar{m}_j), j = 1, \dots, m, \text{ а общая трудоемкость в соответствии с (12) равна:}$$

$$R = \sum_{j=1}^m \bar{R}_j. \text{ Этот показатель должен быть минимизирован за счет числа и состава модулей, т.е.}$$

$$\hat{R} = \min_{n, \Theta_1, \dots, \Theta_n} R.$$

Данную задачу с помощью компьютера можно решить полным перебором, однако, на задачах с количеством блоков, большим 8, время решения становится очень большим. Решение данной задачи аналитическим путем является крайне сложным, поэтому для нахождения псевдо-оптимального решения используются алгоритмы случайного поиска [2].

Выводы

В результате данной работы сформулированы задачи определения оптимальной структуры информационного базиса, получено их решение для случаев равномерной разбивки базиса и его синтеза из заданных блоков. Предложены алгоритмы решения задач определения оптимальной по трудоемкости освоения блочно-модульной структуры информационного базиса для случаев равномерной разбивки базиса и его синтеза из заданных блоков.

Список литературы

1. Андрушко М. И. Модели и методы поддержки принятия информационных решений в интерактивных системах обучения и контроля знаний / М. И. Андрушко, С. Н. Ежеманская Л. Н. Корпачева // Перспективные материалы, технологии, конструкции, экономика: Сб научн. тр. по материалам Всероссийской научно-технической конференции. Вып. 10. – Красноярск: ГУЦМиЗ, 2004. – С. 211-213.
2. Корпачева Л. Н. Алгоритмы оптимизации блочно-модульной структуры информационно-терминологического базиса / Л. Н. Корпачева, С. С. Огнерубов // Вестник университетского комплекса: Сб. науч. трудов/ Под общей ред. профессора Н. В. Василенко. – Вып. 6 (20). – Красноярск: ВСФ РГУИТП, НИИ СУВПТ, 2005. – С. 66-71.
3. Корпачева Л. Н. Интерактивные информационные ресурсы и модели для принятия решений / Л. Н. Корпачева // Перспективные материалы, технологии, конструкции, экономика: Сб. научн. тр. по материалам Всероссийской научно-технической конференции. Вып. 10. – Красноярск: ГУЦМиЗ, 2004. – С. 134-136.
4. Околелов Г. Н. Программированное обучение и новые информационные технологии обучения // Информатика и образование. –1993. – №5. – С.7-19.
5. Тархов С. В. Управление адаптивным обучением и его оптимизация на базе теории абстрактных автоматов и марковских процессов // Информационные технологии моделирования и управления. Научно-технический журнал. – Воронеж: Научная книга, 2005. – №1 (19). – С. 39-45.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

Рецензенты:

Антамошкин Александр Николаевич, д-р техн. наук, профессор, профессор кафедры системного анализа и исследования операций СибГАУ, г. Красноярск.

Ходос Дмитрий Васильевич, д-р экон. наук, доцент, профессор кафедры экономики и информационных технологий менеджмента СФУ, г. Красноярск.

Кирьянов Борис Фёдорович, д.т.н., профессор, профессор кафедры прикладной математики и информатики, ФГБОУ ВПО Новгородский государственный университет им. Ярослава Мудрого, г. Великий Новгород.