УДК 004.94+001:372.82

Т.Н. Дубовик, И.А. Алпатова

КОГНИТИВНЫЕ МОДЕЛИ С ПАРАМЕТРИЧЕСКОЙ АДАПТАЦИЕЙ К ХАРАКТЕРИСТИКАМ ЛИЧНОСТИ. МЕТОДИЧЕСКИЙ АСПЕКТ. ЧАСТЬ 1

Аннотация. Статья посвящена решению актуальной задачи: повышения качества образования путем создания когнитивной модели с параметрической адаптацией. Данная модель представлена в форме регрессионного уравнения, которое связывает между собой показатели усвоения изучаемого материала с характеристиками личности Адаптация модели осуществляется применительно к процессу изучения учебных дисциплин, определенных учебной программой специальности "специализированные компьютерные системы".

Ключевые слова: тренажер, когнитивнаая модель, когорта, кластер, параметрическая адаптация, математическая модель, корреляция, рейтинг.

Постановка задачи

Для исследования влияния комплекса психофизиологических и интеллектуальных факторов на когнитивные процессы используется модель в форме регрессионного уравнения, которое связывает между собой показатели усвоения изучаемого материала (оценки, рейтинги) с характеристиками личности учащегося. К последним относятся такие характеристики как: уровень подготовки по результатам аттестации, интеллектуальные качества, эмоциональная устойчивость, короткая память, долгая память и др. Адаптация модели осуществляется применительно к процессу освоения учебных дисциплин, определенных учебной программой специальности "специализированные компьютерные системы".

Исходными данными служат: перечень изучаемых дисциплин, рейтинги этих дисциплин, перечень факторов, определяющих модель учащегося, тестовые оценки факторов, которые в модели входят в виде переменных, а коэффициенты при этих переменных являются параметрами модели. Эти коэффициенты определяют уровень влияния

© Дубовик Т.Н., Алпатова И.А., 2015

факторов на когнитивные процессы. Данные об оценке этих факторов получаются на основе тестирования учащегося известными тестами [3, 6].

Основная часть

Построение когнитивной модели осуществляется в три этапа.

На первом этапе проводится предварительный анализ парциального влияния каждого из рассмотренных факторов. На основе корреляционного анализа строится матрица влияния факторов на процесс обучения. Для оценки синергетических эффектов вычисляются средние значения коэффициентов корреляции по дисциплинам для каждого фактора и средние значения по факторам для каждой дисциплины.

Средний коэффициент корреляции по дисциплинам вычисляется как среднее арифметическое коэффициентов корреляции рейтингов всех рассматриваемых дисциплин с определенным фактором модели. Этот показатель указывает на относительное влияние данного фактора на усвоение материала по всем рассмотренным дисциплинам. Средний коэффициент корреляции по факторам определяется как среднее арифметическое коэффициентов корреляции всех рассмотренных факторов с одной из дисциплин. Этот коэффициент характеризует степень зависимости данной дисциплины от рассмотренных факторов при их совместном учете в оценках успешности процесса обучения, этот коэффициент может служить мерой синергетического эффекта. С целью определения доминирующих параметров для каждой дисциплины из анализа корреляционной матрицы выделяются те факторы, у которых наблюдается высокий коэффициент парной корреляции с соответствующей дисциплиной, пороговое значение задается (например, $r>0,5$).

На втором этапе для набора выделенных факторов строятся множественные регрессионные уравнения и исчисляются коэффициенты множественной корреляции.

На третьем этапе осуществлялся переход к нормированным переменным для вычисления коэффициентов при соответствующих факторах (параметров модели) для оценки их парциального влияния на качество усвоения материала по различным дисциплинам. Эти коэффициенты и определяют относительный уровень влияния соответствующих факторов модели на качество усвоения материала по изучаемым дисциплинам, которое определяется значениями соответствующих рейтингов.

Описание математической модели

Определим показатель усвоения материала (оценку знаний, рейтинг) y(y_i) как случайную величину, которая выражается через независимую переменную **x**(x_i) где j=1,2…k, k– целое число, x_i - факторы влияния на процесс познания – интеллектуальные способности, психофизиологические факторы и уровень подготовки (определяются тестированием). Рассмотрим линейное регрессионное уравнение с k -переменными. По результатам n выборочных наблюдений отыскиваются коэффициенты (параметры) $\hat{b}_i$ $\left(i=\overline{0,\,k}\right)$ следующего регрессионного уравнения:

$$y=\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\ldots+\hat{b}_kx_k. \qquad (1)$$

Более удобно перейти к нормированным переменным

$$w=\frac{y-m_y}{\sigma_y},\; u_i=\frac{x_i-m_{x_i}}{\sigma_{x_i}},$$

тогда уравнение (1) переходит в уравнение (2)

$$w=\hat{\beta}_1u_1+\hat{\beta}_2u_2+\ldots+\hat{\beta}_ku_k. \qquad (2)$$

Коэффициенты βi в относительных величинах характеризуют степень влияния i-го фактора на результат. Если коэффициенты β_i определять методом наименьших квадратов, то оценки этих коэффициентов находят решением системы уравнений:

$$\left.\begin{array}{l}\hat{\beta}_1 \quad\; + r_{12}\hat{\beta}_2+r_{13}\hat{\beta}_3+\ldots+r_{1k}\hat{\beta}_{1k}=r_{01},\\ r_{21}\hat{\beta}_1+\hat{\beta}_2 \quad\; + r_{23}\hat{\beta}_3+\ldots+r_{2k}\hat{\beta}_k=r_{02},\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ r_{k1}\hat{\beta}_1+r_{k2}\hat{\beta}_2+\ldots\;\ldots\ldots\ldots+\hat{\beta}_k \quad =r_{0k}\end{array}\right\}. \qquad (3)$$

где r_{0i} – оценка коэффициента корреляции между величинами y и x_i, $i=\overline{1,k}$;

r_{ij} – оценка коэффициента корреляции между величинами x_i и x_j, $i,j=\overline{1,k}$.

Из этой системы уравнений следует, что коэффициенты $\hat{\beta}_i$ при коррелированных факторах не могут быть определены независимо друг от друга. Оценки этих коэффициентов могут быть найдены по формулам Крамера:

$$\widehat{\beta}_i = \frac{D_i}{D}, \qquad i = \overline{1,\,k}.$$

Определитель D системы уравнений имеет вид:

$$D = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & 1 & r_{23} & \dots & r_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & r_{k3} & \dots & 1 \end{vmatrix}. \tag{4}$$

Определитель D_i получают из определителя D заменой i-го столбца столбцом правой части системы (3). Оценки коэффициентов b_i находят по формулам:

$$\widehat{b}_i = \widehat{\beta}_i \frac{s_y}{s_{x_i}}, \qquad \widehat{b}_0 = \widehat{y} - \widehat{b}_1 x_1 - \widehat{b}_2 x_2 - \dots - \widehat{b}_k x_k. \tag{5}$$

Средние квадратичные отклонения ошибок коэффициентов $\widehat{b}_0$ и $\widehat{b}_i$ $\left(i = \overline{1,k} \right)$ вычисляются по формулам:

$$s_{\widehat{b}_0} = \frac{\sigma}{\sqrt{n-k-1}}, \quad s_{\widehat{b}_i} = \frac{\sigma}{\sqrt{n-k-1}} \sqrt{\frac{D^*_{ii}}{D^*}}. \tag{6}$$

где D^* – определитель корреляционной матрицы независимых переменных; D^*_{ii} – определитель, получаемый из определителя D^* вычеркиванием i-й строки и i-го столбца. Определитель D^* имеет вид:

$$D^* = \begin{vmatrix} k_{11} & k_{12} & \dots & k_{1k} \\ k_{21} & k_{22} & \dots & k_{2k} \\ \dots & \dots & \dots & \dots \\ k_{k1} & k_{k2} & \dots & k_{kk} \end{vmatrix}, k_{ii} = s^2_{x_i}, \; k_{ij} = r_{x_i x_j} s_{x_i} s_{x_j} \tag{7}$$

В качестве оценки дисперсии σ^2 берем величину s^2. Значение s^2 находится по одной из формул:

$$s^2 = \frac{\sum_{i=1}^{n} \left[y_i - \left(\widehat{b}_0 + \widehat{b}_1 x_{1i} + \widehat{b}_2 x_{2i} + \dots + \widehat{b}_k x_{ki} \right) \right]^2}{n-k-1},$$
$$s^2 = s^2_y \left(1 - R^2 \right) \frac{n-k-1}{n-1} \tag{8}$$

При $n >> k$ параметр s^2 можно определять по формуле:

 ISSN 1562-9945

$$s^2 = s_y^2\left(1 - R^2\right). \qquad (9)$$

При этом коэффициент множественной корреляции R вычисляется по одной из формул:

$$R = \left[\sum_{i=1}^{k} r_{0i}\widehat{\beta}_i\right]^{\frac{1}{2}}, \; R = \left(1 - \frac{s^2}{s_y^2}\right)^{\frac{1}{2}} \qquad (10)$$

$$R = \sum_{i=1}^{n} \frac{\left(y_i - \widehat{y}\right)\left(y_{p_i} - \widehat{y}_p\right)}{\left(n-1\right)s_y s_{y_p}}, \; R = \left(1 - \frac{\Delta}{\Delta_{11}}\right)^{\frac{1}{2}} \qquad (11)$$

где y_i – опытные значения результирующего фактора;

y_{p_i} – значения результирующего фактора, рассчитанные по уравнениям (3).

Определитель Δ представляет собой определитель полной нормированной корреляционной матрицы

$$\Delta = \begin{vmatrix} 1 & r_{01} & r_{02} & \ldots & r_{03} \\ r_{10} & 1 & r_{12} & \ldots & r_{1k} \\ r_{20} & r_{21} & 1 & \ldots & r_{2k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{k0} & r_{k1} & r_{k2} & \ldots & 1 \end{vmatrix}. \qquad (12)$$

Определитель Δ_{11} получают из определителя Δ вычеркиванием первой строки и первого столбца. Этот определитель совпадает с определителем D системы уравнений (3). Скорректированное значение коэффициента множественной корреляции определяется формулой:

$$\widehat{R} = \left[1 - \left(1 - R^2\right)\frac{n-1}{n-k-1}\right]^{\frac{1}{2}}. \qquad (13)$$

Квадрат коэффициента множественной корреляции (коэффициент детерминации) показывает долю дисперсии функции отклика, обусловленную выбранным регрессионным уравнением.

Дисперсия ошибки регрессионного уравнения в точке $\left(x_1^*,\; x_2^*,\; x_3^*, \ldots, x_k^*\right)$ определяется выражением:

$$s_{np}^2 = \left[\frac{\sigma^2}{n-k-1} + \sum_{i=1}^{k}\left(x_i^* - \widehat{x}_i\right)^2 s_{b_i}^2 + \sum_{i>j}\left(x_i^* - \widehat{x}_i\right)\left(x_j^* - \widehat{x}_j\right)K_{b_i b_j}\right] \qquad (14)$$

При этом корреляционный момент между коэффициентами $\widehat{b}_i$ и $\widehat{b}_j$ вычисляется так

$$K_{\widehat{b}_i\widehat{b}_j} = \left(\frac{\sigma^2}{n-k-1}\right)\frac{D_{ij}^*}{D^*} \tag{15}$$

где D^* – определитель корреляционных моментов всех независимых переменных;

D_{ij}^* – алгебраическое дополнение элемента k_{ij}, определитель, получаемый из определителя (4.6) вычеркиванием i-й строки и j-го столбца и умноженный на $(-1)^{i+j}$.

Дисперсия ошибки прогнозирования функции отклика в точке $(x_1^*,\ x_2^*,\dots,x_k^*)$ с учетом разброса индивидуальных значений относительно регрессионного уравнения имеет вид:

$$s_{пр}^2 = \left[\sigma^2 + \frac{\sigma^2}{n-k-1} + \sum_{i=1}^{k}\left(x_i^* - \widehat{x}_i\right)^2 s_{\widehat{b}_i}^2 + 2\sum_{i>j}\left(x_i^* - \widehat{x}_i\right)\left(x_j^* - \widehat{x}_j\right)K_{\widehat{b}_i\widehat{b}_j}\right] \tag{16}$$

В более компактной форме это выражение можно записать так:

$$s_{пр}^2 = \sigma^2\left[1 + \frac{1}{n-k-1} + \frac{1}{n-k-1}\sum_{i,j=1}^{k}\frac{D_{ij}^*}{D^*}\left(x_i^* - \widehat{x}_i\right)\left(x_j^* - \widehat{x}_j\right)\right] \tag{17}$$

Верхняя и нижняя доверительные границы прогнозируемого значения результирующего фактора в точке $(x_1^*, x_2^*,\dots,x_k^*)$ вычисляется по формуле:

$$y_в, y_н = \widehat{b}_0 + \widehat{b}_1 x_1^* + \widehat{b}_2 x_2^* + \ldots + \widehat{b}_k x_k^* \pm t_{\gamma,n-k-1}\, s_{пр} \tag{18}$$

где $t_{\gamma,n-k-1}$ – коэффициент распределения Стьюдента, соответствующий доверительной вероятности γ и числу степенней свободы $\upsilon = n\text{-}k\text{-}1$.

Приведенные выражения можно представить в матричной форме. Уравнение множественной регрессии в матричной форме имеет вид:

$$Y = BX \tag{19}$$

где Y – вектор наблюдений результирующего фактора; X – матрица независимых переменных; В – вектор параметров, подлежащих оцениванию.

 ISSN 1562-9945

Доверительные границы результирующего фактора в точке, определяемой вектором X^*, вычисляются по формуле [1]

$$y_в, y_н = \hat{b}_0 + \hat{b}_1 x_1^* + \ldots + \hat{b}_k x_k^* \pm t_{\gamma, n-k-1}\, s \sqrt{1 + X^{*\prime} \left(X' X \right)^{-1} X^*} \qquad (20)$$

где $t_{\gamma,n-k-1}$ – коэффициент, определяемый из распределения Стьюдента с $(n-k-1)$ степенью свободы; X' – транспонированная матрица наблюдений; X^*– вектор-столбец, характеризующий точку в k-мерном пространстве, в которой производится прогнозирование функции отклика.

Среднее квадратичное отклонение ошибки прогнозирования в точке, определяемой вектором X^*, записывается так:

$$s_{пр} = s \sqrt{1 + X^{*\prime} \left(X' X \right)^{-1} X^*} \qquad (21)$$

Выражения (7), (8) эквивалентны. Однако выражением (7) удобнее пользоваться при определении вклада отдельных составляющих в ошибку прогнозирования оцениваемых параметров [2].

Определения и исходные данные

Введем несколько определений для обобщенных рейтингов. Будем различать следующие показатели:

- рейтинг студента по дисциплине – данные контрольных испытаний из соответствующей ведомости; -рейтинг студента по кластеру (блоку) дисциплин, вычисляется как среднее арифметическое по рейтингам всех дисциплин кластера;

- рейтинг студента по дисциплинам, выбранным по некоторому признаку для целей исследования, вычисляется как среднее арифметическое по рейтингам выбранных дисциплин;

- рейтинг студента по всем изученным на данный момент времени дисциплинам, вычисляется как среднее арифметическое по рейтингам всех изученных дисциплин;

-рейтинг дисциплины, вычисляется как среднее арифметическое по рейтингам всех студентов, изучавших данную дисциплину;

-рейтинг кластера дисциплин, вычисляется как среднее арифметическое по рейтингам дисциплин, включенных в кластер;

-кластеры дисциплин остаются неизменными на протяжении всего исследования, они являются результатом классификации. Кластеры формируются в соответствии с выводами локального феноменологического анализа.

Исходные данные для расчетов формируются в виде двух реляционных баз данных:

1) база данных всех студентов, содержащая рейтинги каждого студента по всем изученным дисциплинам. Таблица размером N1*N2, N1 – число студентов, N2 – число дисциплин.

2) база данных всех студентов, содержащая данные тестов, определяющих интеллектуальные и психофизиологические параметры каждого студента. Таблица размером N1*N3, N1 – число студентов, N3 – число характеристик, определенных набором тестов.

Все дисциплины рабочей программы по специальности "специализированные компьютерные системы" разбиты на три кластера. Первый кластер дисциплин включает цикл дисциплин по программированию, второй – по математическим основам моделирования и алгоритмов, третий кластер включает дисциплины, связанные с созданием технических устройств и элементов компьютерной техники. В каждом из блоков выбираются три условно ведущих (или характерных) дисциплины, которые реализованы в составе компьютерного тренажера.

Для исследований выбираются три когорты студентов – когорта 1 –контроля и когорты 2 и 3 – наблюдения. В когорте 1 контроля (без тренажера, n_1 студентов), в когорте 2 наблюдения (эксперимент с использованием адаптивного тренажера, n_2 студентов), в когорте 3 наблюдения (эксперимент с использованием неадаптивного тренажера, n_3 студентов). Наличие когорты 3 имеет смысл, если имеется набор необходимых статистических данных, это расширяет рамки исследований.

Когорты формируются таким образом, чтобы распределение студентов по успеваемости было статистически близким в смысле близости математических ожиданий и дисперсий рейтингов. Для этих когорт средние значения рейтингов до начала эксперимента обозначены, соответственно, $R_1 \pm r_1$ - когорта1, $R_2 \pm r_2$ – когорта 2, $R_3 \pm r_3$ – когорта 3, где r_1, r_2, r_3 – отклонения значений от среднего. При больших массивах выборок R_1, R_2, R_3 – вычисляются как математические ожидания рейтингов для соответствующих когорт, r_1, r_2, r_3 – вычисляются как среднеквадратичные отклонения, соответственно.

Для эксперимента отбираются по три дисциплины в каждом кластере: с максимальным, средним и минимальным значениями рейтинга, соответственно. Рейтинги 1,2,3 вычисляются для соответствующих

ISSN 1562-9945

групп учащихся (когорт), они преобразовываются в один средний рейтинг для сравнительного анализа рейтингов дисциплин [3,4,5].

В перспективе работы: ранжирование факторов влияния на когнитивные процессы, создание моделей и алгоритма построения когнитивных процессов, формулировка задачи прогноза.

ЛИТЕРАТУРА

1. Кендал М. Многомерный статистический анализ и временные ряды / М. Кендал, А. Стьюарт. – М.: Наука, 1976. – 736 с.
2. Алпатова І. А. Моделі та алгоритми оцінки впливу еколого-гігієнічної ситуації промислового регіону на захворюваність населення хворобами шкіри / І. А. Алпатова // Медична інформатика та інженерія. – 2009. – №2. – С.57–61.
3. Дубовик Т.Н., Сергеева О.В., Дубовик Д.Д. Использование тестовых систем для повышения качества обучения Materiały VIII Międzynarodowej naukowi-praktycznej konferencji «Nauka: teoria i praktyka - 2012» Przemyśl Nauka i studia. 2012 - 88 str. Стр. 31-33.
4. Дубовик Т.М., Курта А, Бєкєдова С. Информационные технологии в управлении сложными системами [текст] / Т.Н. Дубовик // Материалы научной конференции 2011, Днепропетровск, Изд. «Свидлер А.Л.», – С 375 – 377.
5. Дубовик Т.М., Семьонов В.А. К вопросу об адаптации и настройке моделей обучения [текст]/ Т.Н. Дубовик // Региональный межвузовский сборник научных работ «Системные технологии» 3(86) 2013. – С.19–28.
6. Аванесов В.С., Хохолова Т.С., Ступак Ю.А., Потап О.Е., Чернявский В.Г., Плискановський С.А. Педагогические тесты. Вопросы разработки и примменения [Текст] Пособия для проподавателей / В.С. Аванесов, Т.С. Хохолова, Ю.А. Ступак, О.Е. Потап, В.Г. Чернявский, С.А. Плискановський – Дніпропетровськ: Пороги, 2005. – 64 с.