УДК 378.1

Н. Г. Гоголева, Е. А. Яковлева, Е. А. Папичев
Санкт-Петербургский государственный экономический университет

Прогнозирование успеваемости студентов факультета электроники СПбГЭТУ «ЛЭТИ» на основе цепей Маркова

Представлена модель прогнозирования успеваемости студентов на старших курсах на основании данных за первый и второй курсы. Проведены расчеты по представленной модели. Начальными данными для модели является таблица успеваемости студентов шести групп ФЭЛ (факультета электроники) СПбГЭТУ «ЛЭТИ» после первого и второго года обучения (год поступления 2008).

Цепи Маркова, матрица переходов, уравнение регрессии

1. Математическая модель прогнозирования успеваемости студентов с применением цепей Маркова. Процесс называется марковским, если в любой момент времени вероятность любого состояния системы в будущем зависит только от состояния системы в текущий момент и не зависит от того, каким образом система пришла в это состояние [1].

Переходной вероятностью p_{ij} называется условная вероятность перехода системы на k-м шаге в состояние S_j при условии, что на $(k-1)$ шаге система находилась в S_i. Матрицей перехода системы называют матрицу, которая содержит все переходные вероятности этой системы.

Равенство Маркова связывает матрицу перехода P_n за n шагов с матрицей перехода P_1 за 1 шаг:

$$P_n = P_1^n. \qquad (1)$$

За состояние системы примем определенное распределение студентов по группам в зависимости от успеваемости. Считаем, что состояние системы в текущем году зависит только от её состояния в предыдущем году, т. е. определяем данный процесс изменения системы как цепь Маркова.

Система имеет 4 состояния: S_0, S_1, S_2, S_3 – где S_0 – состояние системы после первого года обучения; S_1 – состояние системы после второго года; S_2 – состояние системы после третьего года; S_3 – состояние системы после четвертого года (соответствует окончанию бакалавриата). Используя данные результатов экзаменов после первого курса, вычислим средний балл для каждого студента и разделим всех студентов на n групп. Далее рассмотрим, как меняется количество студентов в каждой группе в процессе обучения.

Состояние системы описывается матрицей-столбцом, каждый элемент которой представляет собой наличие студентов в группе после соответствующего года обучения.

$$S_k = \begin{bmatrix} s_1^k \\ s_2^k \\ \cdot \\ s_n^k \end{bmatrix}.$$

Здесь s_i^k – число студентов, оказавшихся в i-й группе после $(k+1)$-го курса.

Здесь и далее матрицей перехода студентов A_{0k} будем называть матрицу, каждый элемент которой a_{ij} представляет число студентов, перешедших из i-й группы в j-ю после $(k+1)$-го курса.

Элементы матрицы перехода системы P_{0k} определяются по формуле статистической вероятности (статистической вероятностью события называется отношение числа испытаний, в которых проявилось данное событие к общему числу испытаний). Применительно к решаемой задаче формула статистической вероятности имеет вид

$$p_{ij} = \frac{a_{ij}}{s_i^k}, \qquad (2)$$

где a_{ij} – число студентов, перешедших из i-й группы в j-ю после (k + 1)-го курса; s_i^k – число студентов, оказавшихся в i-й группе после (k + 1)-го курса.

Используя собранные данные, получим начальное состояние системы S_0 и матрицу переходов студентов A_{01}:

$$S_0 = \begin{bmatrix} s_1^0 \\ s_2^0 \\ \cdot \\ s_n^0 \end{bmatrix}, \qquad A_{01} = \begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ a_{31} & a_{32} & \cdot & a_{3n} \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix}.$$

Здесь s_i^0 – число студентов, оказавшихся в i-й группе после первого курса; a_{ij} – число студентов, перешедших из i-й группы в j-ю после второго курса.

Для увеличения точности прогноза из модели исключены данные студентов, которых отчислили по причинам, не связанным с академической неуспеваемостью.

Следующим шагом является составление матрицы вероятностей перехода P_{01}, элемент матрицы которой определяется по формуле (2):

$$p_{ij} = \frac{a_{ij}}{s_i^0}. \qquad (3)$$

Далее, используя равенство Маркова, определим матрицу вероятностей переходов из начального состояния S_0 (после первого курса) в состояние S_2 (после третьего курса) за два шага. Используя равенство Маркова (1), получим:

$$P_{02} = P_{01}^2. \qquad (4)$$

Чтобы получить матрицу A_{02} перехода студентов из состояния S_0 (после первого курса) в состояние S_2 (после третьего курса), необходимо матрицу перехода системы P_{02} поэлементно умножить на соответствующие элементы матрицы начального состояния системы, при этом округляя полученные результаты до целого числа. Тогда

$$a_{ij}^{02} = p_{ij}^{02} \cdot s_i^0. \qquad (5)$$

Теперь можно определить число студентов в каждой группе после окончания 3-го курса:

$$s_j^2 = \sum_{i=1}^{n} a_{ij}^{02}. \qquad (6)$$

Полученные значения запишем в матрицу-столбец S_2, каждый элемент которой представляет собой расчетное число студентов в соответствующей группе после окончания 3-го курса. Далее составляем матрицу-столбец $S_2^{'}$, основываясь на данных эксперимента.

Повторяем эту же процедуру для 4-го курса.

Таким образом, имеем реальные данные и спрогнозированные, что позволяет построить уравнение регрессии [2], где зависимой переменной будут реальные данные, а независимой – спрогнозированные.

2. Результаты моделирования. Начальными данными для модели является таблица успеваемости студентов шести групп ФЭЛ (факультета электроники) СПбГЭТУ «ЛЭТИ» после первого и второго года обучения (год поступления 2008).

На основании данных успеваемости студентов после первого курса вычислим средний балл для каждого студента и распределим всех студентов на 11 групп (минимальное количество групп, необходимое для построения уравнения парной линейной регрессии – 7 [2]).

Разделим студентов на 11 групп: 1-я – от 3 до 3,199; 2-я – от 3,2 до 3,399; 3-я – от 3,4 до 3,599; 4-я – от 3,6 до 3,799; 5-я – от 3,8 до 3,999; 6-я – от 4,0 до 4,199; 7-я – от 4,2 до 4,399; 8-я – от 4,4 до 4,599; 9-я – от 4,6 до 4,799; 10-я – от 4,8 до 5,0; 11-я – отчислены.

Используя собранные данные, получим начальное состояние системы S_0 и матрицу перехода A_{01}:

$$S_0 = \begin{bmatrix} 4 \\ 11 \\ 13 \\ 14 \\ 11 \\ 7 \\ 4 \\ 3 \\ 4 \\ 1 \\ 17 \end{bmatrix}; A_{01} = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 4 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 2 \\ 3 & 4 & 0 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 1 \\ 2 & 3 & 3 & 4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 1 & 4 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 17 \end{bmatrix}.$$

Теперь, используя (3), находим матрицу вероятностей перехода P_{01}:

$$P_{01}=\begin{bmatrix} 0,25 & 0,5 & 0 & 0 & 0 & 0 & 0,25 & 0 & 0 & 0 & 0 \\ 0,182 & 0,363 & 0 & 0,091 & 0,182 & 0 & 0 & 0 & 0 & 0 & 0,182 \\ 0,231 & 0,307 & 0 & 0,077 & 0,154 & 0,154 & 0 & 0 & 0 & 0 & 0,077 \\ 0,143 & 0,214 & 0,214 & 0,286 & 0,071 & 0 & 0,071 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,182 & 0,091 & 0,091 & 0,363 & 0,182 & 0,091 & 0 & 0 \\ 0 & 0 & 0 & 0,285 & 0,143 & 0,286 & 0,143 & 0 & 0,143 & 0 & 0 \\ 0 & 0,25 & 0 & 0 & 0 & 0 & 0,25 & 0,25 & 0,25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,667 & 0,333 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,25 & 0 & 0 & 0 & 0,75 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Далее, используя равенство Маркова (4), получаем матрицу перехода P_{02} из начального состояния в состояние 2 за два шага:

$$P_{02}=P_{01}^2=\begin{bmatrix} 0,1535 & 0,369 & 0 & 0,0455 & 0,091 & 0 & 0,125 & 0,0625 & 0,0625 & 0 & 0,091 \\ 0,1246 & 0,2422 & 0,0195 & 0,0922 & 0,0891 & 0,0166 & 0,118 & 0,0331 & 0,0166 & 0 & 0,2481 \\ 0,1246 & 0,2434 & 0,0165 & 0,1219 & 0,0974 & 0,0581 & 0,1411 & 0,028 & 0,036 & 0 & 0,1329 \\ 0,165 & 0,2938 & 0,0612 & 0,1307 & 0,0987 & 0,0394 & 0,0996 & 0,0307 & 0,0242 & 0 & 0,0554 \\ 0,026 & 0,1297 & 0,0389 & 0,0945 & 0,0342 & 0,0571 & 0,1497 & 0,2287 & 0,1727 & 0,0683 & 0 \\ 0,0408 & 0,0967 & 0,061 & 0,189 & 0,0741 & 0,1306 & 0,1488 & 0,0618 & 0,0897 & 0,1072 & 0 \\ 0,0455 & 0,1533 & 0 & 0,0227 & 0,0455 & 0,0625 & 0,0625 & 0,2293 & 0,1457 & 0,1875 & 0,0455 \\ 0 & 0 & 0 & 0 & 0 & 0,0833 & 0 & 0,4449 & 0,2221 & 0,2498 & 0 \\ 0 & 0 & 0 & 0,0712 & 0,0357 & 0,0715 & 0,0357 & 0 & 0,0357 & 0,75 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Используя формулу (5), составляем A_{02}. Теперь, используя (6), можно определить число студентов в каждой группе после окончания 3 курса. Полученные данные запишем в матрицу-столбец S_2. Далее составляем матрицу-столбец S_2', основываясь на данных эксперимента:

$$A_{02}=\begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 3 \\ 2 & 3 & 0 & 2 & 1 & 1 & 2 & 0 & 0 & 0 & 2 \\ 2 & 4 & 1 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 2 & 3 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 17 \end{bmatrix},\ S_2=\begin{bmatrix} 6 \\ 14 \\ 1 \\ 7 \\ 4 \\ 5 \\ 9 \\ 7 \\ 5 \\ 8 \\ 23 \end{bmatrix},\ S_2'=\begin{bmatrix} 0 \\ 11 \\ 11 \\ 4 \\ 8 \\ 8 \\ 5 \\ 9 \\ 3 \\ 3 \\ 27 \end{bmatrix}.$$

Аналогичные вычисления проделываем для 4-го курса, получаем матрицу перехода P_{03} из начального состояния в состояние 3 за три шага и составляем A_{03}. Затем определяем число студентов в каждой группе после окончания 4-го курса и записываем полученные данные в матрицу-столбец S_3. Далее составляем матрицу-столбец S_3', основываясь на данных эксперимента.

$$S_3 = \begin{bmatrix} 6 \\ 15 \\ 0 \\ 4 \\ 4 \\ 4 \\ 6 \\ 6 \\ 7 \\ 11 \\ 26 \end{bmatrix}, \quad S_3^{'} = \begin{bmatrix} 1 \\ 4 \\ 5 \\ 2 \\ 7 \\ 7 \\ 8 \\ 11 \\ 7 \\ 8 \\ 29 \end{bmatrix}.$$

Сравнение экспериментальных и расчетных данных распределения студентов по 11 группам представлено на рис. 1 и 2. На рис. 1 сравниваются теоретические и экспериментальные результаты успеваемости после окончания 3-го курса. На рис. 2 сравниваются теоретические и экспериментальные результаты успеваемости после окончания 4-го курса. По горизонтальной оси отложен номер группы i, по вертикальной оси – число студентов в данной группе (s_i^2 – число студентов в i-й группе после окончания 3-го курса;

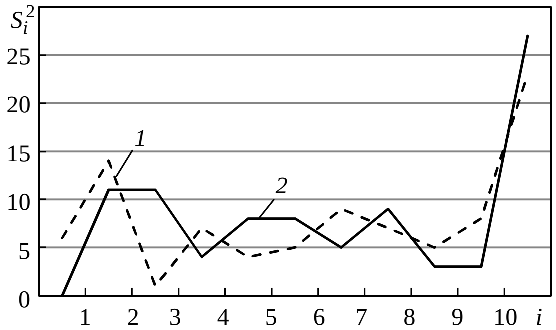

Рис. 1

s_i^3 – число студентов в i-й группе после окончания 4-го курса). Линией 1 обозначены теоретические результаты, полученные с использованием уравнения Маркова. Линией 2 обозначены экспериментальные данные, взятые из таблицы успеваемости студентов.

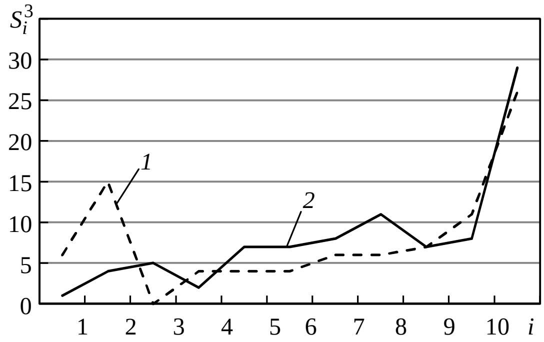

Рис. 2

Далее были построены регрессионные модели успеваемости студентов после 3-го и 4-го курсов. Составлялись уравнения парной линейной регрессии $y = a + b \cdot x$ [2], где переменной x обозначены реальные данные, а переменной y – спрогнозированные. Результаты сведены в таблицу.

Параметр	3-й курс	4-й курс
a	0,876166	1,468436
b	0,89171	0,818508
Коэффициент корреляции ($r(xy)$)	0,733289	0,776728
Коэффициент детерминации (R^2)	0,537713	0,603306

Видно, что коэффициент корреляции для обоих случаев больше 0.7 (коэффициент детерминации больше 0.5), что говорит о сильной связи между исследуемыми параметрами (один признак определяет другой больше, чем на половину).

3. Представлена методология оценки и прогнозирования качества образования на примере СПбГЭТУ «ЛЭТИ». Эта методология основана на аппарате Марковских цепей. Моделирование показало удовлетворительное совпадение результатов расчета с экспериментальными данными. Чтобы сделать окончательный вывод о возможности применения данной методики, необходимы дальнейшие исследования.

СПИСОК ЛИТЕРАТУРЫ

1. Письменный Д. Конспект лекций по теории вероятностей, математической статистике и случайным процессам. М.: Айрис-Пресс, 2008.

2. Кремер Н. Ш., Путко Б. А. Эконометрика. М.: ЮНИТИ, 2008.

N. G. Gogoleva, E. A. Yakovleva, E. A. Papichev
Saint Petersburg State University of Economic

STUDENTS GRADES PROGNOSTICATING BY MARKOV CHAINS («LETI», FACULTY OF ELECTRONICS)

The mathematical model of students grades prognosticating, based on Markov chains, is developed. The students grades at elder courses are predicted on the base of the one at 1-2 courses. The calculation results are presented. The initial data is the table of grades of six groups of students (Saint Petersburg Electrotechnical University "LETI" , Faculty of Electronics).

Markov Chains, transition matrix, regression equation