

УДК 621.3

DOI:10.25729/ESI.2025.38.2.014

Прогностическая модель оценки успешности обучения на основе метрической идентификации индивидуального образовательного маршрута

Степанов Михаил Александрович¹, Гергет Ольга Михайловна²

¹Томский политехнический университет «ИШИТР»,

Россия, Томск, *mas65@tpu.ru*

²Институт проблем управления им. В.А. Трапезникова РАН,

Россия, Москва

Аннотация. В статье предложена комплексная модель оценки успешности обучения на основе методов теории распознавания и обработки многомерных данных в рамках деятельностного направления, которое включает широкий спектр современных методик обучения, основанных на интеллектуальном анализе данных обучающегося, его цифрового следа, планирования элементов в составе учебно-методических комплексов и прогнозировании результатов обучения. Траектория обучения формируется, как поэлементная кластеризация взаимосвязанных последовательностей данных, в частности, трудоемкости элементов дисциплин, социально-демографических параметров обучающихся, результатов промежуточных аттестаций освоения рабочих программ. Реализована нейросетевая модель прогнозной экстраполяции результатов обучения. Представлены результаты экспериментальных исследований модели прогнозирования успешности обучения по выборкам данных, аккумулированным в цифровых системах управления обучением.

Ключевые слова: индивидуальная образовательная траектория, пространственная экстраполяция, параметрическая идентификация, прогноз успешности обучения, многослойная нейронная сеть, метрические методы, метод Кульбака-Лейблера

Цитирование: Степанов М.А. Прогностическая модель оценки успешности обучения на основе метрической идентификации индивидуального образовательного маршрута / М.А. Степанов, О.М. Гергет // Информационные и математические технологии в науке и управлении, 2025. – № 2(38). – С. 170-181. – DOI:10.25729/ESI.2025.38.2.014.

1. Введение. Модернизация российского образования в соответствии с Федеральной программой развития образования [1] на ближайшие годы нацелена на глубокую трансформацию существующей системы образования Российской Федерации, что определяет значительную потребность в использовании новых аналитико-информационных инструментов для эффективного управления образовательным процессом на разных уровнях – государственном, вузовском, локальном. Одной из характерных особенностей учебного процесса является информатизация образования, основанная на внедрении новых информационных технологий, ориентированных на оптимальное управление ходом учебного процесса во всех его проявлениях: аудиторной работе студентов, курсовом и дипломном проектировании, самостоятельной работе, научно-исследовательской работе, учебно-методическом обеспечении предметов, составлении расписаний занятий, сбора и анализа информации об успеваемости студентов и уровня преподавания, кадрово-финансовом и административном аспектах деятельности образовательного учреждения [2, 3]. Вместе с тем, современная образовательная практика высшей школы имеет ограниченные возможности поиска и проявления субъектами образовательного процесса собственной индивидуальности, что является следствием противоречия между групповой формой организации обучения по специальностям в вузе и индивидуальным характером усвоения знаний студентом. Алгоритмизированный режим функционирования образовательных институтов предполагает выполнение равных правил, указаний, инструкций, в том числе при попытках алгоритмизировать построение индивидуальных траекторий обучения при помощи

математических методов [4-10]. В этих условиях реализация индивидуальных образовательных траекторий отождествляется с правом студента выбирать образовательные компоненты в ограниченном объеме, что недостаточно для обеспечения условий его полноценного профессионального саморазвития. Одним из направлений преодоления указанных трудностей является разработка современных методик и техник с применением интеллектуальных компьютерных технологий мониторинга результативности, таких, как построение и отслеживание индивидуальной траектории обучения или внедрение игровых технологий в учебный процесс на основе теории управления и интеллектуального анализа данных с применением машинного обучения и математической статистики. Целью исследований является разработка математического представления индивидуального образовательного маршрута обучающегося, на основании которого появляется возможность осуществить достоверный прогноз успешности обучения студента. При таком подходе модель прогноза оперирует не только такими параметрами, как успеваемость и трудоемкость элементов дисциплины, но и включает социально-демографические параметры обучающихся.

2. Формализация и общий алгоритм формирования индивидуального образовательного маршрута обучающегося. В рамках настоящих исследований индивидуальная образовательная траектория Ω_s^N s -го студента ($s=\overline{1, M}$, где M – общее число студентов выборки) на интервале времени всего обучения по учебному плану специализации в ВУЗе $[t_0, t_2]$ представляется математическим объектом, содержащим N независимых параметров. N -й параметр, регистрируемый в момент времени t_2 , характеризует результат итоговой аттестации в соответствии с r -м учебным планом специализации ($r=\overline{1, R}$, где R – общее количество специализаций в ВУЗе).

Параметры индивидуальной образовательной траектории: $(N-1)$, $(N-2)$, 1 регистрируются на интервале реализации учебного плана $[t_0, t_1]$, где $t_1 < t_2$, и представляют собой интегральную параметризованную оценку степени успешности прохождения мероприятий промежуточного контроля, предусмотренного рабочими программами и тематическими планами дисциплин, составляющих r -й учебный план.

Индивидуализация образовательного процесса предполагает, что для каждого обучающегося возможно изменение количества часов на определенные темы дисциплин, при сохранении параметров и структуры учебного плана. Глубиной девиации образовательного процесса в настоящем исследовании принята трудоемкость одной дисциплины. При этом, принцип индивидуализации будет соблюден и в случае варьирования трудоемкости тем, модулей и других элементов рабочих программ или учебного плана. В апостериорном случае, т.е. после завершения обучения, траектория является полностью определенной в каждой точке контрольных мероприятий.

Принцип, лежащий в основе технологии индивидуальной образовательной траектории заключается в параметрической идентификации недоопределенного математического объекта Ω_s (т.е. полученного по данным на интервале $[t_0, t_1]$, где $t_1 < t_2$) на основе множества типовых образовательных траекторий $\{\Omega\}$ с известной степенью успешности завершения образовательного процесса и определенных на всем временном интервале реализации учебного плана $[t_0, t_2]$. Математический объект – многомерная кривая (многомерная, благодаря множеству одновременно регистрируемых параметров учебного процесса). Например, пока студент наблюдается, до его текущего курса математический объект определен, но нам нужно спрогнозировать успешность прохождения дисциплин и аттестации в будущем – там он не доопределен.

Так как смысл технологии индивидуальной образовательной траектории состоит не только в апостериорном распознавании индивидуального маршрута каждого студента, но и в управлении им с целью коррекции и вывода на заданный уровень успешности прохождения итоговой аттестации, то идентификация образовательного маршрута реализуется по его начальному фрагменту в момент t_1 (рисунок 1). Например, по окончании 2-3 семестров наблюдается, к какой группе траекторий принадлежит образец на своем интервале 2-3 семестров, соответственно, к какому итогу по итоговой аттестации приведет его обучение без коррекции. Если прогнозируемый итог не устраивает, то, начиная с момента времени t_1 , необходимо корректировать траекторию за счет изменения последующих трудоемкостей элементов рабочих программ и учебного плана в направлениях, приближающих к необходимой (эталонной) траектории с необходимой степенью успешности итоговой аттестации.

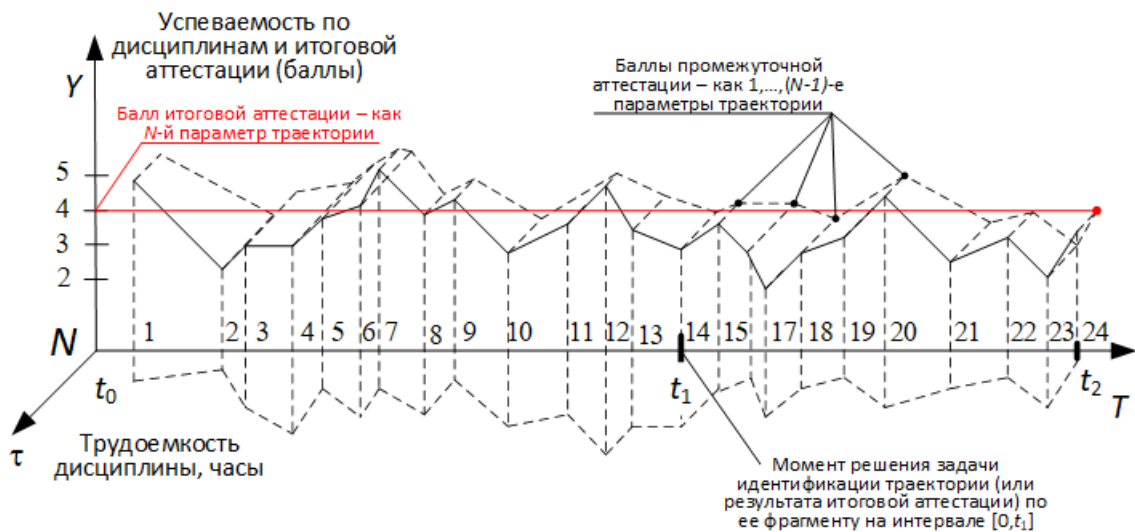


Рис. 1. Геометрическая интерпретация апостериорной образовательной траектории Ω размером $N=24$ в трехмерном пространстве «Дискретное время семестровых контрольных мероприятий-Успеваемость-Трудоемкость»

Коррекция индивидуальной образовательной траектории может осуществляться за счет изменения соотношения часов аудиторных занятий и самостоятельного изучения темы, изменения тематических планов в части практических, лабораторных или лекционных занятий, при сохранении общей трудоемкости дисциплины, а также дисциплин по выбору. Коррекция может быть осуществлена путем приведения оставшейся траектории к типовому эталону заданной степени успешности и специализации траектории, а сами эталоны формируются по выборке данных прошедших обучения студентов.

Таким образом, основные компоненты и операции по формированию индивидуального образовательного маршрута обучающегося включают:

1. Трехмерное пространство $\{X, Y, Z\}$ построения траектории, где ось X – ось времени реализации учебного плана специализации в дискретной шкале отчетных мероприятий по элементам рабочих программ, ось Y – порядковая целочисленная шкала степени успешности прохождения отчетных мероприятий по элементам рабочих программ, ось Z – шкала трудоемкости элементов рабочих программ.
2. Собственно, траектория Ω_s^N s -го студента, представляющая собой вектор: $\Omega_s^N = [y_1, y_2, \dots, y_N]^T$, где $y_i, i=1, \dots, N$ – значение степени успешности прохождения отчетных мероприятий по элементам рабочих программ в момент времени t_i .

Математический объект Ω_s^N представим в виде вектора данных, а при $N \rightarrow \infty$ – кривую в трехмерном пространстве или семейство двух плоских кривых.

3. Для решения задачи идентификации – как решения задачи сличения с известным, необходимо иметь эталонные кривые, представляющие собой алфавит классов различных итогов обучения (или различных степеней успешности обучения по r -му учебному плану специализации в ВУЗе). В свою очередь, для получения таких эталонов необходимо решить задачу кластеризации объектов Ω_s^N в пространстве $\mathbf{R}^{M \times N}$, где M – общее число студентов выборки) на интервале времени всего обучения по учебному плану специализации в ВУЗе $[t_0, t_2]$.

4. Метрика сличения фрагментов кривых траекторий – в общем случае Минковского, с

варьируемыми переменными p и q : $l_p^q(X, A) = \left(\sum_{j=1}^N |x_j - a_j|^p \right)^{\frac{q}{p}}$, где X – сличаемый образец – фрагмент траектории, A – эталон траектории.

5. Δt_i – коррекция, вычисляемая в часах, которые необходимо добавить или отнять у каждого i -го элемент учебного плана или рабочей программы на интервале $[t_1, t_2]$ для перевода обучающегося на требуемую (эталонную) траекторию.

Общий алгоритм построения и коррекции индивидуальной траектории студента на основе игровых технологий будет выглядеть следующим образом (рисунок 2).

Как видно из рисунка 2, центральным элементом алгоритма является этап идентификации фрагмента траектории, как промежуточных достижений и результатов студента для оценки соответствия траектории заданным параметрам. Достоверность решения задачи идентификации – как отнесения к заданному, зависит от трех составляющих задачи – статистической представительности эталонов (шаблонов) и сличаемых образцов, словаря признаков и алфавита классов.

3. Модель метрической идентификации индивидуального образовательного маршрута обучающегося. Предположим, что по результатам прохождения за время t_1 цикла или модуля, посвященного изучению определённой темы, студент получил оценку y_1 баллов. Тогда образовательная траектория студента за период реализации учебного плана $[t_0, t_1]$ будет представлять собой одну реализацию выборки, подлежащей кластеризации, с целью эталонирования траекторий $\Omega_s^N = [y_1, y_2, \dots, y_N]^T$, где $y_i, i=1, \dots, N$ – степень успешности прохождения игры в части реализации элементов рабочих программ в момент времени t_i .

Цифровые системы управления обучением (LMS) в ВУЗах содержат сведения по студентам гораздо более широкие, чем дискретное время семестровых контрольных мероприятий, текущая успеваемость студента и трудоемкость пройденных дисциплин. Например, в датасетах [11-13] успеваемости студентов в качестве основания для выработки прогноза успешности итоговой аттестации, совместно с данными по рубежным контрольным точкам могут быть использованы такие параметры, как: 1) посещаемость занятий; 2) количество часов работы по неполному рабочему дню студента; 3) количество часов участия во внеклассных мероприятиях; 4) количество часов самостоятельного обучения каждую неделю, и др.; 5) мало вариабельные во времени социо-демографические данные студента; 6) пол (гендер); 7) национальность; 8) размер семьи; 9) состояние здоровья; 10) уровень образования родителей; 11) параметризация рода занятий отца и матери; 12) семейное положение; 13) параметризация карьерных устремлений студента, и др.

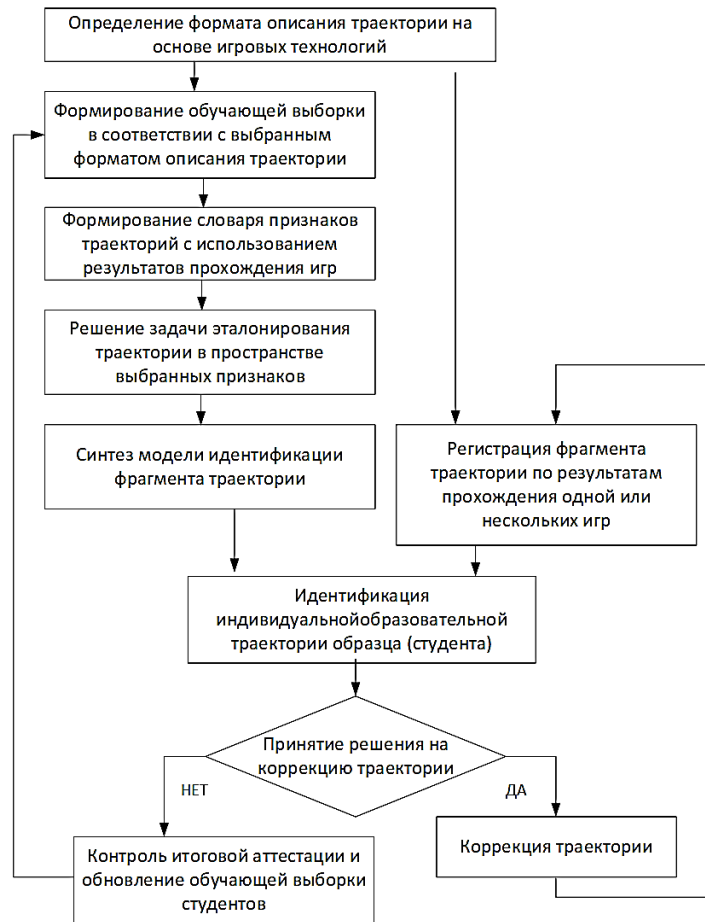


Рис. 2. Этапы алгоритма построения индивидуальной траектории студента

Наличие такого количества данных по каждому студенту обуславливает возможность осуществлять прогноз успешности обучения на основе метрической идентификации индивидуального образовательного маршрута не только во времени, но и в пространстве вышеперечисленных, преимущественно неизменных на протяжении всего обучения, характеристик студента (рисунок 3).

Дата	здор.	сем. полож	состав семьи	род. занят.	национа льн.	происхо жд.	раб. день	внеу рочн	сам. подг	Посещ.	Дисц.1	Дисц.2	Дисц.N	ИМЭ	Диплом
08.01.2022							28,07	0,39	64,14	65,26	13,25				
09.01.2022							27,86	0,12	64,89	62,16	13,74				
09.01.2022							28,06	0,20	63,86	55,25	13,74				
12.01.2022							27,78	0,18	65,65	56,01	13,90				
13.01.2022							28,21	0,30	64,74	56,00	13,78				
14.01.2022							28,31	0,16	64,40	56,39	13,43				
16.01.2022							28,24	0,18	64,33	58,04	13,32				
17.01.2022							28,13	0,22	64,67	58,32	13,97	15,22			
18.01.2022							31,90	0,13	67,99	53,01	13,55	14,94			
03.05.2022							28,35	0,41	65,64	58,07	12,96	14,60			
04.05.2022							28,57	0,35	65,63	57,92		14,52			
05.05.2022							28,49	0,40	65,76	56,96		14,47			
06.05.2022							28,40	0,44	65,79	56,97		14,30			
07.05.2022							31,31	0,38	64,81	57,07					
08.05.2022							28,86	0,35	63,97	56,81				15,65	
09.05.2022							26,83	0,35	64,08	57,08				15,69	
10.05.2022							27,09	0,40	64,13	56,64				15,95	
11.05.2022							30,11	0,34	64,60	60,34				15,92	
12.05.2022							29,98	0,36	64,56	59,90				16,74	
13.05.2022							30,07	0,34	64,55	60,08				16,68	
03.06.2022							50,60	0,49	70,19	64,50				16,32	
05.06.2022							50,64	0,32	70,04	62,31				16,46	
		28,49	0,40	65,76	56,96	0,30	64,74			пространственная экстраполяция				4,00	5,00

Рис. 3. Пример топологии исходных данных для реализации индивидуального и группового прогнозирования во времени (освоения элементов учебного плана) и пространстве социо-демографических данных студента (в таблице представлены оценки и баллы)

В результате появляется возможность одновременного применения и комплексирования инструментария, как для распознавания многомерных кривых, построенных по значениям временных отсчетов параметризованных событий учебного плана, так и метрической идентификации в пространстве слабо-вариабельных во времени социо-демографических данных студента.

В случае, если временных отсчетов мероприятий учебного плана в периоде основания прогноза оказывается слишком много или они сильно зашумлены, то альтернативой является замена временного ряда на его статистические моменты, т.е. вторичные признаки.

Протокол наблюдений Ω_1 прецедентов обучающих векторов вида в таком случае будет представлять собой матрицу обучающей выборки вида «временные и пространственные (или вторичные) признаки – класс успешности итоговой аттестации» (1):

$$\Omega_1 = \begin{bmatrix} k_1^1 & \dots & k_{L_k}^1 & v_1^1 & \dots & v_{L_v}^1 & y_1^1 & \dots & y_m^1 & p_1^1 & \dots & p_{L_p}^1 & z_1^1 & \dots & z_{L_z}^1 & q_1^1 & \dots & q_N^1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ k_1^i & \dots & k_{L_k}^i & v_1^i & \dots & v_{L_v}^i & y_1^i & \dots & y_m^i & p_1^i & \dots & p_{L_p}^i & z_1^i & \dots & z_{L_z}^i & q_1^i & \dots & q_N^i \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ k_1^L & \dots & k_{L_k}^L & v_1^L & \dots & v_{L_v}^L & y_1^L & \dots & y_m^L & p_1^L & \dots & p_{L_p}^L & z_1^L & \dots & z_{L_z}^L & q_1^L & \dots & q_N^L \end{bmatrix}, \quad (1)$$

где $(n+L_z+N)$ – размер параметрического описания каждого прецедента; L – количество строк матрицы, характеризует объем обучающей выборки; k – параметризованные данные о прогрессе изучения дисциплины студентом; v – переменные промежуточной аттестации; y – переменные посещаемости; p – социально-демографические данные; z – переменные состояния обучающегося (здоровье, психологический скрининг и т.д.); q – переменные итоговой аттестации – прогнозируемые целевые переменные.

Таким образом, прогностическая модель оценки успешности обучения на основе метрической идентификации индивидуального образовательного маршрута включает в себя несколько инструментов экстраполяции (пространственной и временной) в одно общее множество прогнозируемых переменных – баллы мероприятий итоговой аттестации, с последующей мажоритарной обработкой результатов.

В алгоритме комплексирования использовались следующие механизмы вычисления принадлежности классу успешности итоговой аттестации:

- метрические, по критериям минимума а) евклидоваго расстояния, б) расстояний Чебышёва и Минковского третьей степени;
- статистический на основе рассогласования Кульбака-Лейблера;
- на основе максимума нормированной функции кросскорреляции;
- на основе многослойной нейронной сети (НС) прямого распространения с векторной функцией выхода.

Типовая многослойная НС прямого распространения сигнала имеет архитектуру $\{L, X_1, \dots, X_N, 1\}$, где L – размер входного слоя, X_1, \dots, X_N – размеры скрытых слоев или $\{ML, X_1, \dots, X_N, 1\}$ – при распознавании по M характеристикам учебного процесса (рисунок 4) [14-16].

Идентифицируемая кривая многомерна за счет большого числа регистрируемых параметров учебного процесса, а также характеристик самого студента.

Применение НС (рисунок 4) для обнаружения и классификации фрагмента многомерной кривой, принадлежащей эталонной траектории студента, осуществлялось в соответствии с формулой:

$$y_j = f_j^{[3]} \left(\sum_{m=1}^{n_2} w_{mj}^{[3]} \left(f_m^{[2]} \left(\sum_{h=1}^{n_1} w_{hm}^{[2]} \left(f_h^{[1]} \left(\sum_{i=1}^n w_{ih}^{[1]} x_i \right) \right) \right) \right) \right), \quad j = 1, N, \quad (2)$$

где f – передаточная функция нейронов, w – синаптические коэффициенты нейронной сети, N – число нейронов выходного слоя, кодирующего степень успешность обучения.

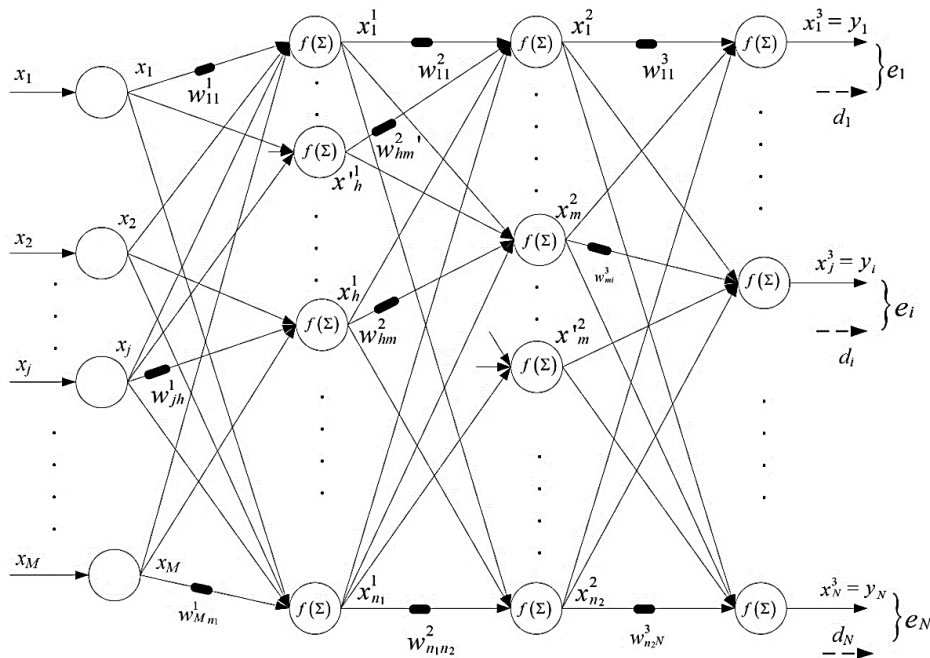


Рис. 4. Типовая многослойная НС для временной экстраполяции успешности итоговой аттестации

Время на рисунке в неявном виде – каждое значение x_j – это оценка по каким-либо мероприятиям промежуточной аттестации или результатам прохождения игр. Это все процессы – во времени.

Если представить в N -мерном пространстве признаков два сличаемых фрагмента в виде двух точек X и A , то *евклидово расстояние* между ними вычисляется, как

$$l_2(X, A) = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + \dots + (x_n - a_n)^2} = \sqrt{\sum_{i=1}^N (x_i - a_i)^2}. \quad (3)$$

Если евклидово расстояние возвести в квадрат, то при классификации назначаются большие веса более отдаленным друг от друга объектам.

При $\nu \rightarrow \infty$ в формуле $l_\nu^\mu(X, A) = \left(\sum_{j=1}^N |x_j - a_j|^\nu \right)^{\frac{\mu}{\nu}}$, расстояние (по Чебышёву) равно

максимальному отклонению по какой-либо из координат $l_\infty = \max_{1 \leq j \leq N} |x_j - a_j|$, т.е. выбор данной меры расстояния с большой степенью дает резкое выделение класса с наименьшим расстоянием до кривой.

Для оценки степени линейной зависимости двух кривых X и A использовался коэффициент корреляции:

$$r_{xa} = \mu_{xa} / \sigma_x \sigma_a, \quad (4)$$

где σ_x^2 , σ_a^2 – дисперсии, а μ_{xa} – их ковариация (совместная дисперсия). При наличии k наблюдений в каждом фрагменте

$$\mu(X, A) = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})(a_i - \bar{a}). \quad (5)$$

В качестве нормированной корреляционной функции применялась:

$$p_I(t_1, t_2) = \frac{K_I(t_1, t_2)}{\sigma_I(t_1)\sigma_I(t_2)} \quad \text{или} \quad p_I(t_1, t_2) = \frac{K_I(t_1, t_2)}{\sqrt{K_I(t_1, t_1)}\sqrt{K_I(t_2, t_2)}}, \quad (6)$$

где I – значения отсчетов кривой образца, J – значения отсчетов кривой эталона.

При обнаружении совпадений двух траекторий взаимная корреляционная функция для двух фрагментов кривых – окна сканируемой последовательности x и шаблона a как коэффициент корреляции между x_t и a_{t+k} в зависимости от номера отсчета k вычислялся параметр:

$$r_k = \frac{\sum_{t=1}^{n-k} X_t A_{t+k} - \sum_{t=1}^{n-k} A_t \sum_{t=1}^{n-k} X_{t+k} / (n-k)}{\sqrt{\left[\sum_{t=k}^{n-k} A_t^2 - \sum_{t=k}^{n-k} A_t^2 / (n-k) \right] \left[\sum_{t=k+1}^n X_t^2 - \sum_{t=k+1}^n X_t^2 / (n-k) \right]}} \quad (7)$$

Максимальное значение параметра (7) указывает на искомую эталонную реализацию траектории, соответствующую определенному результату на итоговой аттестации.

Статистические методы обнаружения, как и большинство методов распознавания, основаны на гипотезе компактности, согласно которой фрагменты кривых одного класса (образца) более близки друг другу, чем фрагменты кривых разных классов. Расстояние (информационное расхождение, дивергенция – является безразмерной величиной) Кульбака-Лейблера – это метрика между распределениями случайных величин, или расстояние $D_{KL}(P||Q)$ от статистического распределения P значений кривой образца до статистического распределения Q значений кривой шаблона. В непрерывном случае:

$$D_{KL}(P||Q) = \int p(x) \ln(p(x)/q(x)) dx, \quad (8)$$

для дискретного случая:

$$D_{KL}(P||Q) = \sum_{i=1}^n P_i \ln(P_i/Q_i), \quad (9)$$

где P_i, Q_i – отсчеты значений функции распределения двух кривых.

Так как рассогласование Кульбака-Лейблера между шаблоном и скользящим окном d_{kl}^1 в общем случае не равно рассогласованию между скользящим окном и шаблоном d_{kl}^2 , то для его использования в качестве расстояния применяется показатель $d_{KL} = (d_{kl}^1 + d_{kl}^2)/2$.

4. Экспериментальные исследования. В качестве датасета, примененного для экспериментальных исследований модели прогнозирования успешности обучения, использовалась база данных [12]. Данные содержат 33 параметра, из которых 33-й «G3» – прогнозируемый – результат обучения по математике студентов португальского ВУЗа. 32 и 31-й, «G1» и «G2» – результаты промежуточной аттестации. Датасет содержит сведения по 395 студентам. Таким образом данные позволяют осуществлять прогноз успеваемости как пространственной (по параметрам с 1-го по 30-й), так и временной (в периоде основания параметры – «G1» и «G2») экстраполяцией (рисунок 5). Чем G3 ближе к 20, тем успешнее обучение.

```
school;sex;age;address;famsize;Pstatus;Medu;Fedu;Mjob;Fjob;reason;guardian;traveltime;studytime;failures;schoolsup;famsup;paid;activities;nursery;higher2;internet;romantic;famrel;freetime;goout;Dalc;Walc;health;absences;G1;G2;G3
"GP";"F";18;"U";"GT3";"A";4;4;"at_home";"teacher";"course";"mother";2;2;0;"yes";"no";"no";"no";"yes";"yes";"no";"no";4;3;4;1;1;3;6;"5";"6";6
"GP";"F";17;"U";"GT3";"T";1;1;"at_home";"other";"course";"father";1;2;0;"no";"yes";"no";"no";"yes";"yes";"no";5;3;3;1;1;3;4;"5";"5";6
"GP";"F";15;"U";"LE3";"T";1;1;"at_home";"other";"other";"mother";1;2;3;"yes";"no";"yes";"no";"yes";"yes";"yes";"no";4;3;2;2;3;3;10;"7";"8";10
"GP";"F";15;"U";"GT3";"T";4;2;"health";"services";"home";"mother";1;3;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"yes";3;2;2;1;1;5;2;"15";"14";15
"GP";"F";16;"U";"GT3";"T";3;3;"other";"other";"home";"father";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"no";"no";4;3;2;1;2;5;4;"6";"10";10
"GP";"M";16;"U";"LE3";"T";4;3;"services";"other";"reputation";"mother";1;2;0;"no";"yes";"yes";"yes";"yes";"yes";"yes";"no";5;4;2;1;2;5;10;"15";"15";15
"GP";"M";16;"U";"LE3";"T";2;2;"other";"other";"home";"mother";1;2;0;"no";"no";"no";"no";"yes";"yes";"no";4;4;4;1;1;3;0;"12";"12";11
"GP";"F";17;"U";"GT3";"A";4;4;"other";"teacher";"home";"mother";2;2;0;"yes";"yes";"no";"no";"yes";"yes";"no";"no";4;1;4;1;1;6;"6";"5";6
"GP";"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;2;2;1;1;0;"16";"18";19
... ..
```

Рис. 5. Фрагмент датасета «Student Performance» <https://archive.ics.uci.edu/dataset/320/student+performance>

Целью экспериментов было определить наилучшие метрики и модели, позволяющие получить наиболее достоверный прогноз, а также оценить выигрыш от их комплексирования.

На рисунке 6 представлены результаты распознавания степени успешности итоговой аттестации относительно заранее сформированных эталонных траекторий при различных объемах исходных данных и используемых моделей прогнозирования.

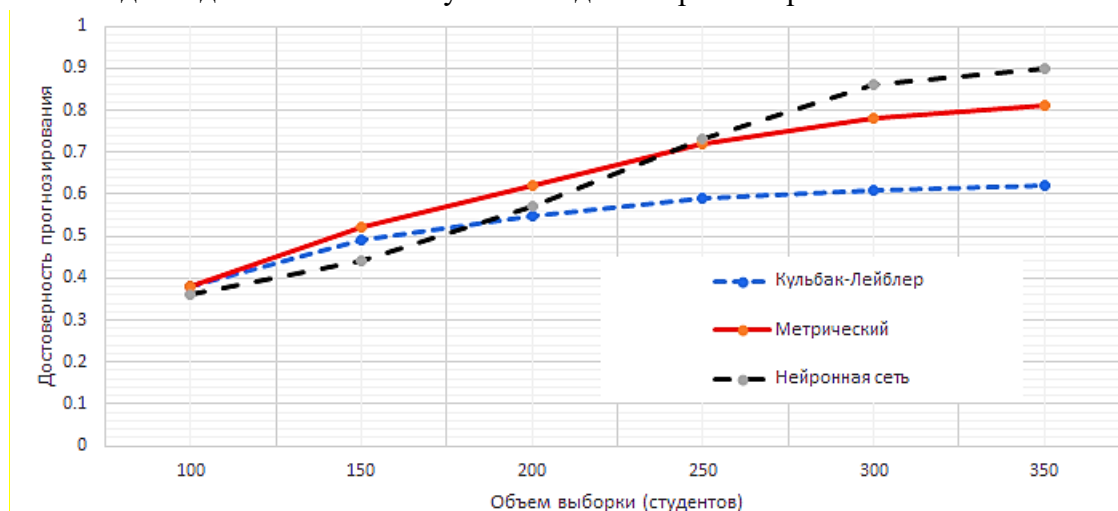


Рис. 6. Зависимость достоверности прогнозирования тремя методами параметра успешности обучения (G33) от объема обучающей выборки

Распознанная траектория однозначно указывала на эталонный прогноз успешности обучения по математике в 3-м (заключительном) семестре. При использовании НС, выборка была поделена на обучающую (от 100 до 350 студентов) и проверочную – 45 студентов.

При использовании метрических (в детерминированной постановке) и статистических моделей для эталонирования кривых образовательных траекторий использовалось различное число студентов. Достоверность прогнозирования рассчитывалась частотным способом:

$$D = M/N, \quad (10)$$

где M - число успешных прогнозов, N – общее число опытов.

Каждый опыт отличался случайным составом примеров обучающей выборки и новым обучением нейронной сети. Низкая достоверность метода на основе сличения функций распределения Кульбака-Лейблера обусловлена зашумленностью данных, т.к. некоторые переменные в датасете сформированы по результатам субъективных оценок при анкетировании. Увеличение достоверности нейросетевого метода можно объяснить учетом нелинейностей при накоплении примеров в данных, а снижение метрического метода (относительного нейросетевого) – погрешностью, обусловленной эталонированием большого числа примеров при сличении с шаблоном.

5. Заключение. В результате проведенных исследований предложено представлять индивидуальную образовательную траекторию в виде математического объекта – многомерного массива ретроспективных и слабо-девиантных данных каждого студента, позволяющего осуществлять как индивидуальное, так и групповое прогнозирование, как во времени – по данным промежуточной аттестации и характеристикам освоенных элементов учебного плана, так и в пространстве социально-демографических данных обучающихся. Такое представление траектории позволяет применить классические метрические (в детерминированной постановке) и статистические методы распознавания фрагментов кривых, эталоны которых однозначно указывают на вид успешности итоговой аттестации студента. Среди экспериментально исследованных методов прогнозирования максимальную достоверность показал метод на основе трехслойной нейронной сети, что позволяет предположить наличие в исходных данных сложных взаимозависимостей между

переменными и незамкнутых областей классов успешности обучения в многомерном пространстве разнородных признаков образовательной траектории.

Максимальную достоверность прогноза показывают корреляционные и нейросетевые методы. При этом нейросетевые модели демонстрируют схожую корреляционной точность распознавания взаимозависимостей между переменными при увеличении объема обучающей выборки.

Список источников

1. Государственная программа Российской Федерации «Развитие образования». – URL: <https://docs.edu.gov.ru/document/id/3998> (дата обращения 17.07.2024).
2. Криони Н.К. Инноватика и инновационные образовательные технологии: монография / Н.К. Криони, С.Г. Селиванов, Ф.В. Шарипов. – М.: Редакционно-издательский дом Российского нового университета, 2020. – 296 с.
3. Инновационные образовательные технологии и методы обучения: монография / С.Н. Викжанович, Т.Ю. Четверикова, Е.А. Романова и др. – Saint-Louis, MO: Publishing house science and innovation center, 2014. – 162 с.
4. Носков М.В. Прогностическая модель оценки успешности предметного обучения в условиях цифровизации образования / М.В. Носков, Ю.В. Вайнштейн, М.В. Сомова и др. // Вестник Российского университета дружбы народов. Серия: Информатизация образования, 2023. – Т. 20. – № 1. – С. 7-19. – DOI: 10.22363/2312-8631-2023-20-1-7-19.
5. Алдунин Д.А. Математическая модель для построения оптимальной индивидуальной образовательной траектории обучающегося при изучении массовых открытых онлайн-курсов / Д.А. Алдунин, Г.Г. Федин // Информационные технологии, 2019. – Том 25. – № 4. – С. 250-256. – DOI: 10.17587/it.25.250-256.
6. Rózewski P., Zaikin O. Integrated mathematical model of competence-based learning-teaching process. Bulletin of the Polish academy of sciences: technical sciences, 2015, no. 1 (63), pp. 245-259, DOI: 10.1515/bpasts-2015-0029.
7. Lin J., Pu H., Li Y. et al. Intelligent recommendation system for course selection in smart education. Procedia computer science, 2018, vol. 129, pp. 449-453, DOI: 10.1016/j.procs.2018.03.023.
8. Kabathova J., Drlik M. Towards predicting student's dropout in university courses using different machine learning techniques. Applied sciences, 2021, vol. 11, no. 7, pp. 137-146, DOI: 10.3390/app11073130.
9. Riestra-González M., del Puerto Paule-Ruiz M., Ortin F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & education, 2021, vol. 163, pp. 104-108, DOI: 10.1016/j.compedu.2020.104108.
10. Помян С.В. Прогноз результатов успеваемости студентов вуза на основе марковских процессов / С.В. Помян, О.С. Белоконь // Вестник Вятского государственного университета, 2020. – № 4 (138). – С. 63-73. – DOI: 10.25730/VSU.7606.20.057.
11. Cortez P., Silva A.M.G. Using data mining to predict secondary school student performance. Published in proceedings of 5th annual future business technology conference (2022), 2008.
12. Student performance. Available at: <https://archive.ics.uci.edu/dataset/320/student+performance> (accessed: 11/04/2024).
13. Hasan R. Dataset of student's performance using student information system, moodle and mobile application 'eDify', 2021, DOI:10.5281/zenodo.5591907.
14. Ростовцев В.С. Искусственные нейронные сети. 4-е изд. / В.С. Ростовцев. – СПб.: Лань, 2024. – 216 с.
15. Бишоп К.М. Распознавание образов и машинное обучение / К. М. Бишоп. – СПб.: ООО «Диалектика», 2020. – 960 с.
16. Аггарвал Ч. Нейронные сети и глубокое обучение / Ч. Аггарвал. – СПб.: ООО «Диалектика», 2020. – 752 с.

Гергет Ольга Михайловна. Доцент, доктор технических наук, Институт проблем управления им. В.А. Трапезникова РАН, ведущий научный сотрудник. AuthorID: 115286, SPIN: 9352-4215, ORCID: 0000-0002-6242-9502, gerget@ipu.ru. 117997, Россия, г. Москва, ул. Профсоюзная, 65,

Степанов Михаил Александрович. Томский политехнический университет, Инженерная школа информационных технологий и робототехники. AuthorID: 167502, mas65@tpu.ru, 634050, Россия, г. Томск, ул. Ленина, 30.

UDC 621.3

DOI:10.25729/ESI.2025.38.2.014

A predictive model for assessing learning success based on metric identification of an individual educational route

Mikhail A. Stepanov¹, Olga M. Gerget²

¹Tomsk polytechnic university “ISHITR”,
Russia, Tomsk, *mas65@tpu.ru*

²V.A. Trapeznikov institute of control sciences,
Russia, Moscow

Abstract. The article proposes a comprehensive model for assessing the success of learning based on the methods of recognition theory and processing of multidimensional data within the activity-based direction, which includes a wide range of modern teaching methods based on intelligent analysis of student data, his digital footprint, planning elements in educational and methodological complexes and predicting learning outcomes. The learning trajectory is formed as an element-by-element clustering of interconnected sequences of labor intensity, intermediate results of mastering work programs for a large sample of students. A neural network model for predictive extrapolation of learning outcomes has been developed. The results of experimental studies of the model on data samples accumulated in digital learning management systems are presented.

Keywords: individual educational trajectory, spatial extrapolation, parametric identification, learning success prediction, multilayer neural network, metric methods, Kullback-Leibler method

References

1. Gosudarstvennaya programma Rossiyskoy Federatsii “Razvitiye obrazovaniya” [State program of the Russian Federation "Education Development"]. Available at: <https://docs.edu.gov.ru/document/id/3998> (accessed: 07/17/2024).
2. Krioni, N.K., Selivanov S.G., Sharipov F.V. Innovatika i innovatsionnyye obrazovatel'nyye tekhnologii: monografiya [Innovation and innovative educational technologies: monograph]. M.: Redaktsionno-izdatel'skiy dom Rossiyskogo novogo universiteta [Moscow, editorial and publishing house of the Russian new university], 2020, 296 p.
3. Vikzhanovich S.N., Chetverikova T.Yu., Romanova E.A. et al. Innovatsionnyye obrazovatel'nyye tekhnologii i metody obucheniya: monografiya [Innovative educational technologies and teaching methods: monograph]. Saint-Louis, MO: Publishing house science and innovation center, 2014, 162 p.
4. Noskov M.V., Vainshtein Yu.V., Somov M.V. et al. Prognosticheskaya model' otsenki uspekhov predmetnogo obucheniya v usloviyakh tsifrovizatsii obrazovaniya [Predictive model for assessing the success of subject learning in the context of digitalization of education]. Vestnik Rossiyskogo universiteta druzhby narodov, Seriya: Informatizatsiya obrazovaniya [Bulletin of peoples' friendship university of Russia, Series: informatization of education], 2023, vol. 20, no. 1, pp. 7-19, DOI: 10.22363/2312-8631-2023-20-1-7-19.
5. Aldunin D.A., Fedin G.G. Matematicheskaya model' dlya postroyeniya optimal'noy individual'noy obrazovatel'noy trayektorii obuchayushchegosya pri izuchenii massovykh otkrytykh onlayn-kursov [Mathematical model for constructing an optimal individual educational trajectory of a student when studying massive open online courses]. Informatsionnyye tekhnologii [Information technologies], 2019, vol. 25, no. 4, pp. 250-256, DOI: 10.17587/it.25.250-256..
6. Rózewski P., Zaikin O. Integrated mathematical model of competence-based learning-teaching process. Bulletin of the Polish academy of sciences: technical sciences, 2015, no. 1 (63), pp. 245-259, DOI: 10.1515/bpasts-2015-0029.
7. Lin J., Pu H., Li Y. et al. Intelligent recommendation system for course selection in smart education. Procedia computer science, 2018, vol. 129, pp. 449-453, DOI: 10.1016/j.procs.2018.03.023.
8. Kabathova J., Drlik M. Towards predicting student's dropout in university courses using different machine learning techniques. Applied sciences, 2021, vol. 11, no. 7, pp. 137-146, DOI: 10.3390/app11073130.
9. Riestra-González M., del Puerto Paule-Ruiz M., Ortin F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & education, 2021, vol. 163, pp. 104-108, DOI: 10.1016/j.compedu.2020.104108.
10. Pomyan S.V., Belokon O.S. Prognoz rezul'tatov uspevayemosti studentov vuza na osnove markovskikh protsessov [Prediction of academic performance of university students based on markov processes]. Vestnik Vyatskogo

- gosudarstvennogo universiteta [Bulletin of Vyatka state university], 2020, no. 4 (138), pp. 63-73, 10.25730/VSU.7606.20.057.
11. Cortez P., Silva A.M.G. Using data mining to predict secondary school student performance. Published in proceedings of 5th annual future business technology conference (2022), 2008.
 12. Student performance. Available at: <https://archive.ics.uci.edu/dataset/320/student+performance> (accessed: 11/04/2024).
 13. Hasan R. Dataset of student's performance using student information system, moodle and mobile application 'eDify', 2021, DOI:10.5281/zenodo.5591907.
 14. Rostovtsev V.S. *Iskusstvennyye neyronnyye seti. 4-ye izd.* [Artificial neural networks. 4th ed.]. SPb.: Lan' [SPb.: Lan], 2024, 216 p.
 15. Bishop K.M. *Raspoznavaniye obrazov i mashinnoye obucheniye* [Pattern recognition and machine learning]. SPb.: OOO "Dialektika" [St. Petersburg: OOO "Dialectics"], 2020, 960 p.
 16. Aggarwal C. *Neyronnyye seti i glubokoye obucheniye* [Neural networks and deep learning]. SPb.: OOO "Dialektika" [St. Petersburg: OOO "Dialectics"], 2020, 752 p.

Gerget Olga Mikhailovna. Associate professor, doctor of technical sciences, V.A. Trapeznikov institute of control sciences of the Russian Academy of Sciences, Leading Researcher. AuthorID: 115286, SPIN: 9352-4215, ORCID: 0000-0002-6242-9502, gerget@ipu.ru, 117997, Russia, Moscow, Profsoyuznaya St., 65.

Stepanov Mikhail Aleksandrovich. Tomsk Polytechnic University, School of information technology and robotics. AuthorID: 167502, mas65@tpu.ru, 634050, Russia, Tomsk, Lenin St., 30.

Статья поступила в редакцию 07.02.2025; одобрена после рецензирования 06.03.2025; принята к публикации 02.05.2025.

The article was submitted 02/07/2025; approved after reviewing 03/06/2025; accepted for publication 05/02/2025.