

Искусственный интеллект в моделях регионального управления социально-экономическими процессами

Троценко А. Н.

Научно-исследовательский институт Социальных Систем при МГУ имени М. В. Ломоносова, Москва, Российская Федерация; info@niiss.ru

РЕФЕРАТ

В работе анализируются методология создания информационно-технологической платформы «цифровых портретов» сложных социально-экономических систем на примере региональных систем России как одного из необходимых элементов стратегического управления процессами достижения национальных целей.

Для решения задач, связанных с созданием цифровых портретов, определены практические способы применения технологий Data Mining, алгоритмов искусственного интеллекта, методов математической статистики, линейной алгебры, а также кластерного и экономического факторного анализа. На этой основе предложены модели для целей цифровой трансформации региональных систем управления социально-экономическими процессами.

Ключевые слова: национальные цели развития, цифровая трансформация, интеллектуальный анализ данных, кластерный регрессионный анализ, математические модели обеспечения управленческих решений.

Для цитирования: Троценко А. Н. Искусственный интеллект в моделях регионального управления социально-экономическими процессами // Управленческое консультирование. 2026. № 1. С. 99–117. EDN VCDVUF

Artificial Intelligence in Models for Regional Management of Socio-Economic Processes

Anatoly N. Trotsenko

Research Institute for Social Systems at M.V. Lomonosov Moscow State University, Moscow, Russian Federation; info@niiss.ru

ABSTRACT

This paper analyzes methodological approaches for creating an info-technological platform for “digital portraits” of complex socioeconomic systems using the example of regional systems in Russia, as a necessary element of strategic management of the national development goals achievement.

To address the challenges associated with creating digital portraits, practical applications of Data mining technologies, artificial intelligence algorithms, mathematical statistics, linear algebra, and cluster and economic factor analysis are identified. Based on these approaches, models are proposed for the digital transformation of regional socioeconomic management systems.

Keywords: development goals, digital transformation, Data Mining, cluster regression analysis, mathematical models for supporting management decisions.

For citation: Trotsenko A. N. Artificial Intelligence in Models for Regional Management of Socio-Economic Processes // Administrative Consulting. 2026. N 1. P. 99–117. EDN VCDVUF

Введение

Эффективное решение задач управления сложными социально-экономическими системами должно изначально учитывать актуальные результаты и выводы их анализа на основе методологии теории социального выбора. Эти результаты

указывают на то, что на уровне национальной экономики и социальной сферы России определение приоритетов и стратегических ориентиров развития преимущественно отнесено к сфере ответственности политической системы стран¹. В качестве одного из примеров этого положения могут служить цели национального развития, определенные Указом Президента Российской Федерации от 7 мая 2024 года № 309 (далее — Указ)², которые формируют содержание ключевых направлений стратегического развития на ближайшие десятилетия. При этом цели социального развития (укрепление здоровья населения, реализация потенциала и талантов человека, создание комфортной, безопасной и экологичной среды) должны быть обеспечены устойчивой и динамичной *технологической трансформацией национальной экономики*, причем с акцентом на опережающее технологическое развитие регионов [21]. Иными словами, сформулированные направления можно рассматривать в качестве целевых ориентиров для построения новой модели социально-экономического развития страны, основанной на поиске и реализации достижимых и подкрепленных ресурсами *стратегий социально-экономического развития* всей совокупности субъектов РФ, включающих формирование общих концепций обеспечения устойчивости базовых потребностей общества, а также возможностей активации точек роста и опережающего развития региональных экономик и всей страны в целом [22].

Следует отметить соответствующую потребность в реализации существенных, но сбалансированных институциональных преобразований («структурных реформ») всех систем управления [21], которые предполагают необходимость применения новых подходов к реализации стратегических, координационных и стабилизационных функций государства на основе *цифровой трансформации* всей многоуровневой системы государственного управления. Иными словами, цифровая трансформация государственного управления является основной технологической предпосылкой структурных реформ, необходимых для эффективного достижения указанных целей стратегического развития экономики и социальной сферы. Практическим результатом этого процесса должна стать единая цифровая платформа — система мониторинга, прогнозирования и оценки эффективности достижения целей развития на всех уровнях государственного администрирования. При этом ключевым условием успешной реализации этой платформы должно стать создание и практическое внедрение так называемых цифровых портретов (ЦП) региональных систем субъектов РФ [10].

Очевидно, что эффективное *цифровое управление* сложными социально-экономическими системами (в частности, региональными системами) сопряжено с необходимостью обработки больших объемов данных, достаточно разнородных по содержанию, полноте, наличию явных и скрытых взаимных связей, а также по источникам и методам их сбора. Поэтому помимо использования теоретических междисциплинарных знаний, связанных с государственным администрированием, создание и использование цифровых портретов предполагает применение алгоритмов количественного анализа, основанных на концепции конвергенции естественного и искусственного (машинного) интеллекта [2; 3], а также на общих принципах и методологии, обычно обозначаемой понятием Data Mining (DM-методологии), т. е. совокупности средств и методов приобретения новых знаний на основе интеллектуального анализа больших данных [7; 12; 26; 36]. По сути, DM-методология определяется как совокупность способов организации информационно-технологических

¹ Некипелов А. Д. О природе социального выбора // РСМ. 2006. № 4. С. 5–15. URL: <https://cyberleninka.ru/article/n/o-prirode-sotsialnogo-vybora> (дата обращения: 27.11.2025).

² Указ Президента Российской Федерации «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года» от 7 мая 2024 года № 309. [Электронный ресурс]. URL: <http://publication.pravo.gov.ru/document/0001202405070015>

процессов и применения математических методов, нацеленных на обнаружение в больших массивах данных ранее неизвестных, но практически полезных и доступных для интерпретации знаний, в частности, необходимых для принятия оптимальных решений в сфере экономического управления. Эти знания должны, в том числе, обеспечивать описание связей между количественными показателями, характеризующими анализируемый объект (в данном случае — региональную систему), т. е. предсказывать изменения одних показателей за счет изменения других. В зависимости от целей и требований к решению различных прикладных задач, математическая составляющая Data Mining, как правило, связана с применением междисциплинарного комплекса различных методов. При этом этот комплекс в основном ассоциируется с понятием *искусственного интеллекта* как совокупности *исполняемых в автоматизированном режиме* алгоритмов, основанных на инструментарию математической статистики, технологиях искусственных нейронных сетей (ИНС) и машинного обучения, а также на ряде приложений линейной алгебры, дискретной математики, теории групп и математической логики [13].

Практическое применение принципов Data Mining при разработке и реализации цифровых платформ управления сложными объектами всегда подразумевает реализацию этапной, внутренне связанной схемы последовательного решения определенных блоков задач. Эта схема должна последовательно детализировать следующие ключевые аспекты построения ЦП:

- содержательный анализ, классификация и необходимые преобразования исходных данных в целях идентификации и категоризации конкретных процессов в региональных социально-экономических системах, включая исследования и выявление значимых взаимных связей между показателями, характеризующими их состояние и динамику, а также оценку их состоятельности;

- формализация количественных зависимостей индикаторов состояния и динамики системы от ее управляемых показателей (выбор оптимальных алгоритмических подходов и построение математических моделей описания и управления процессами), а также оценка уровня достоверности модельных оценок и анализ их погрешностей;

- создание логико-математической модели и адаптивного интерфейса для принятия управленческих решений на основе концепции конвергенции «человек — искусственный интеллект».

Исходя из этого, общая задача разработки цифровых портретов региональных социально-экономических систем должна рассматриваться именно как подобная многостадийная схема необходимых процедур, где каждая стадия разработки должна использовать наиболее эффективные способы соответствующего решения. С учетом этого контекста предметом настоящей работы является анализ возможных информационно-технологических и математических подходов, в том числе — алгоритмов искусственного интеллекта, с точки зрения их потенциального эффективного применения на различных этапах решения общей задачи построения ЦП регионов.

Теоретические основы методологии построения цифровых платформ управления

Прежде чем обратиться к выбору наиболее эффективных способов решения задач в рамках указанной схемы создания ЦП, следует сделать необходимые общие уточнения, связанные с термином «искусственный интеллект» (ИИ).

Современная многовариантность реализаций и стремительное расширение сфер применения технологий искусственных нейронных сетей (ИНС) [30] явились причиной достаточно распространенной, но некорректной трактовки терминов ИИ и ИНС как тождественных понятий. На самом деле, в универсальном смысле ИИ — это

область научных знаний, связанная с реализацией *различных* математических моделей, способных *в автономном режиме* решать широкий спектр задач, требующих человеческого интеллекта, в то время как ИНС — это лишь часть подобных моделей, основанная на современных представлениях о работе нейронов в человеческом мозге и, соответственно, на математической имитации этой работы. Кроме того, практическая применимость ИНС часто фундаментально ограничена спецификой требований и исходных условий при создании цифровых платформ, подобных предмету настоящей работы [4; 33; 35]. В частности, применение подобных моделей требует наличия больших и репрезентативных наборов исходных данных (обучающих, контрольных и тестирующих выборок) по каждому показателю анализируемого объекта управления. Однако в случае создания ЦП объем данных по *отдельным показателям* для каждого конкретного региона существенно ограничен ввиду среднегодового способа представления региональной статистики [27], а также реальным горизонтом актуальности данных, характеризующих относительно стабильную динамику региональной системы во времени [10]. С учетом этого содержательно значимые для анализа временные ряды региональных данных обычно не превышают 10–12 элементов выборки, т. е. в нотациях математической статистики имеют не более 8–10 степеней свободы [15]. Это критически влияет на возможности формирования обучающих и контрольных выборок в алгоритмах ИНС. Кроме того, применение ИНС сопряжено с известными проблемами, связанными со строгими требованиями *интерпретируемости результатов* моделирования процессов при использовании ЦП в рамках *систем регионального управления* [4; 33]. Поэтому оба ограничения, по сути, исключают содержательно адекватное применение этого класса алгоритмов ИИ. С другой стороны, общий объем региональной статистики весьма обширен (за счет множества учитываемых показателей для десятков регионов) и, следовательно, требует привлечения различных, рассматриваемых далее, способов многокомпонентного анализа исходных данных [13], основанных на различных алгоритмах ИИ, но не связанных с технологией ИНС [26].

1. Идентификация процессов управления на основе анализа исходных данных статистики

На этом этапе создания ЦП регионов применяются преимущественно экспертные, теоретико-качественные методы анализа [17]. Массив исходной социально-экономической статистики [27] изначально структурируется в виде наборов данных двух типов: показателей, содержательно ассоциированных с характеристиками стратегических целей развития (далее — индикаторы), и показателей, потенциально влияющих на динамику этих индикаторов (далее — факторы), в том числе — факторов, управляемых на уровне региона (далее — управляющие факторы), а также факторов, связанных с «опосредованным» влиянием на динамику региональных процессов (далее — дополнительные факторы) [11]. Очевидно, что подобный анализ может быть основан исключительно на экспертных междисциплинарных исследованиях привлекаемых фокус-групп специалистов, связанных с региональной тематикой, поскольку он едва ли может быть связан с применением математических классификационных алгоритмов, например, основанных на цифровых таксономических технологиях [12]. Предварительные итоги подобного анализа показывают, что в контексте установленных ориентиров развития субъектов РФ может быть выделено до нескольких десятков связанных с национальными целями индикаторов и порядка сотни потенциально влияющих на их изменения факторов. Исходя из этого, на следующем шаге этого этапа формируется совокупность из порядка нескольких сот «элементарных» процессов (т. е. возможных корреляционных связей каждого из индикаторов как минимум с одним из факторов [10]). Эта совокупность «элементарных» процессов является основой для следующего этапа создания ЦП

регионов — применения количественных методов анализа, связанных с установлением как наличия корреляций в парах индикатор — фактор, так и с оценкой их причинно-следственного характера [16; 19; 28].

2. Оценки состоятельности «элементарных» процессов управления

Одним из предварительных условий реализации этого этапа анализа является необходимость *численной гармонизации* индикаторов и особенно факторов, имеющих различную природу и, соответственно, выраженных в исходной статистике в разных числовых размерностях. Чтобы исключить проблемы подобного рода, используется *индексное представление* индикаторов и управляющих факторов, т. е. переменные в паре индикатор — фактор выражаются в нормированных, взаимно сопоставимых по масштабу единицах. Кроме того, применение подобной *индексной формализации* является эффективным способом для объективного сравнения оцениваемых характеристик в разных регионах, а также для корректного учета их изменений во времени [10; 21]. В этой связи в различных экономических приложениях существует множество методических подходов к индексному представлению переменных [1; 6; 14; 20; 21], среди которых оптимальным для рассматриваемого случая представляется аналитическая индексация [6], где индексы должны, прежде всего, обладать ясным содержательным смыслом и обеспечивать интерпретируемость оценок влияния индекс-факторов на изменения индекс-индикаторов в соответствующих процессах [5]. Применяются три взаимно согласованные типа формализации индексов [10; 11]:

$$I_x = M \frac{X(t)}{(P(t) \times Q(t))}, \quad (1)$$

$$I(X_1, X_2) = M \frac{X_1(t)}{X_2(t)}, \quad (2)$$

$$I(X_1) = M \frac{X_1(t)}{X_1(t_0)}. \quad (3)$$

В выражении (1): $X(t)$ — исходный целевой показатель в естественной размерности; $P(t)$ — исходный дополнительный фактор, связанный с временной коррекцией значений X (например, коррекцией на инфляционные эффекты), $Q(t)$ — исходный дополнительный фактор для получения удельных величин (например, численность населения), t — указатель на год сбора статистики [27], а M — масштабированный коэффициент, обеспечивающий соразмерность исходных показателей. Соответственно, в выражении (2): X_1 и X_2 — содержательно связанные показатели исходной статистики, где X_1 по смыслу является частью (долей) показателя X_2 , или X_1 и X_2 — характеристики разных, но содержательно связанных процессов (например, коэффициенты рождаемости и смертности [27]). Наконец, в выражении (3): $X_1(t_0)$ — значение X_1 , соответствующее базовому (опорному) моменту времени (году) сбора статистики t_0 [11; 27]. Кроме того, в отдельных случаях значения числителя в (1)–(3) могут иметь более сложное содержание, когда целесообразно учитывать комбинацию несколько исходных показателей статистики, например, с помощью выражения вида:

$$X_1 = \alpha Q_1 + \beta Q_2 + \dots,$$

где Q_1, Q_2 — показатели исходной статистики размерности X_1 , а α, β — долевые безразмерные коэффициенты, также представленные в исходных данных статистики [27].

При этом, независимо от типа индексации (1)–(3), на отдельных последующих этапах создания ЦП может применяться так называемая стандартная форма индексов [5; 20; 28; 29]:

$$\delta I_i = \frac{I_i - I_m}{\bar{I}}; \quad I_m = \frac{\sum_i^N I_i}{N}, \quad (4)$$

$$\delta I_i = \frac{I_i - I_m}{\sigma_I}; \quad \sigma_I = \sqrt{\frac{\sum_i^N (I_i - I_m)^2}{N-1}}, \quad (5)$$

где, δI_i — стандартизованное представление относительного изменения индекса I_i для i -го компонента анализируемой выборки размером из N элементов, I_m — выборочное среднее (математическое ожидание) индекса I , а σ_I — несмещенная оценка среднеквадратического отклонения индекса I в выборке размером N [15].

С учетом приведенных индексных представлений может быть построена методика оценки состоятельности «элементарных» процессов, основанная на применении класса алгоритмов ИИ, известных как классификация с помощью фиксированных правил [26; 33; 35]. На этом этапе предполагается линейная зависимость пар индикаторов и факторов, причем при анализе для всех пар пока не учитывается потенциальная корреляция между факторами. Отметим, что даже с учетом вынужденного использования малых выборок данных, для гипотезы о линейной зависимости в паре индикатора (Y) и фактора (X) имеются существенные основания. Поскольку в этом случае условное (выборочное) распределение *случайной величины* Y на практике неизвестно, то применяется стандартная для эконометрики процедура *сглаживания экспериментальных данных* [28]. Например, для оценки зависимого от X математического ожидания случайной величины индикатора Y , или $M_x(Y) = f(X)$, выбирается линейная функция в предположении, что пара случайных величин (X , Y) имеет совместное нормальное распределение [16]. При этом, даже если в паре (X , Y) нет индивидуального или совместного нормального распределения, то существуют известные, содержательно эквивалентные преобразования (в частности, функционалы, используемые в корреляционном анализе пары переменных (X , Y)), которыеобладают близким к нормальному распределением. Примером может служить Z -преобразование Фишера, применяемое в оценках значимости линейного коэффициента корреляции R_{XY} между X и Y :

$$Z(R_{XY}) = \frac{1}{2} \ln \frac{1+R_{XY}}{1-R_{XY}}; \quad \sigma_Z = \frac{1}{\sqrt{N-3}}; \quad Z_{\text{расч.}} = \frac{z(R_{XY})}{\sigma_Z} > 2, \quad (6)$$

где σ_Z — среднеквадратическое отклонение величины Z в рамках выборки размера N (оцениваемое с помощью формул типа (5)), а $Z_{\text{расч.}}$ — соответствующий показатель значимости линейного коэффициента корреляции R_{XY} , оцененного в пространстве координат (X , Y) [8]. Таким образом, даже в этом конкретном случае зависимость $Y = f(X)$ с достаточными основаниями может быть представлена в виде линейной функции:

$$Y = f(X) = a + bX + \varepsilon, \quad (7)$$

где оценки параметров зависимости a , b и ε определяются на основе известных методов регрессионного анализа [15; 16; 28]. Дополнительным аргументом в пользу выбора вида зависимости (7) является тот факт, что линейная регрессионная модель, в сравнении, например, со степенным или экспоненциальным случаем,

обеспечивает наименьший риск существенной ошибки прогноза вне поля корреляции данных исходной статистики [19].

Также важно, что первый шаг методики (т. е. наличие корреляции в паре (X, Y)) может быть построен с учетом упомянутых эффектов горизонта актуальности данных исходной статистики [9; 10]. Основные детали метода учета этих эффектов состоят в следующем.

Для каждой пары (X, Y) из общей, отсортированной по годам доступной выборки размера $N_{\text{общ}}$ формируются частичные выборки значений от наиболее актуальных к более ранним годам сбора исходных данных. Далее в рамках каждой частичной выборки рассчитываются коэффициенты корреляции $R_{XY}(df)$, где $df = N - 2$ — число степеней свободы выборки с частичным размером N , а затем — на основе универсальных табличных величин критерия Стьюдента ($\tau_{\text{таб.}}$) определяются соответствующие (для данного df) критические значения линейного коэффициента корреляции — $R_{\text{кр.}}(\tau_{\text{таб.}}(df, \alpha))$, где α — установленный уровень значимости (обычно принимается $\alpha = 0,05$) [8]. Исходя из этого выбирается оптимальный размер частичной выборки N_{opt} , т. е. объективный горизонт актуальности исходных данных для конкретной пары (X, Y) .

Таким образом, эта стадия рассматриваемой классификационной методики ИИ реализуется на основе анализа следующего состава взаимосвязанных правил [26; 33; 35]³:

- значение $R_{XY}(N_{\text{opt}})$ должно быть статистически значимым [15], т. е. должно выполняться соотношение $R_{XY}(N_{\text{opt}}) > R_{\text{кр.}}$;
- должно соблюдаться эквивалентное условие (6), т. е. размер N_{opt} не может быть меньше 5 элементов;
- при соблюдении первых двух условий оптимальный размер N_{opt} выбирается в соответствии с максимально допустимым значением df .

На заключительном шаге оценки состоятельности «элементарных» процессов выполняется автоматизированный анализ линейных регрессионных зависимостей выбранных «коррелирующих» пар индикатор — фактор [11]. Для этого, с учетом вида зависимости (7) с помощью метода наименьших квадратов (МНК) для каждой реализации пары случайных величин (X, Y) строятся *формальные оценки* параметров a , b , ϵ , т. е. свободного члена A , коэффициента регрессии — B и остаточного члена e соответственно [28]. Далее выполняется количественный анализ оценок отдельных параметров и всей однокомпонентной регрессии в целом на соответствие требованиям *несмещенности, состоятельности и эффективности* [8]. Для этого вычисляется классический набор соответствующих критериев, таких как критерии Стьюдента для коэффициента корреляции $R_{XY}(N_{\text{opt}})$, а также оценок A и B , коэффициент детерминации $(R_{XY})^2$, критерий Фишера $F((R_{XY})^2)$, средний модуль ошибки аппроксимации $M(|e_i|)$ и средний коэффициент эластичности $E(B)$ [25]. В рамках этой стадии процедуры также производится автоматизированный анализ величин и характера изменений регрессионных остатков e_i , $i = 1, N_{\text{ит}}$, с точки зрения их соответствия стандартным требованиям к эффективности линейных регрессий (в том числе — несмещенности среднего значения e : $M(e_i) = 0$, независимости e^2 от X , гомоскедастичности и отсутствия автокорреляции в значениях e_i), известных как условия Гаусса — Маркова [32]. Кроме того, производится оценка параметров интервального прогноза на основе построенных парных регрессий, в частности, математического ожидания прогнозных значений Y и соответствующих доверительных интервалов [19], на основе которых формируются правила определения характера

³ По сути, эта процедура реализует один из вариантов классификационного алгоритма ИИ с помощью метода решающих деревьев, одного из вариантов процедуры на основе фиксированных правил.

динамики (тренда) процесса. В свою очередь, приемлемость всех расчетных величин указанных критериев оценивается путем сравнения с их соответствующими критическими значениями из универсальных таблиц по распределениям Стьюдента и Фишера (для конкретных df и уровня значимости α) [28].

Кратко описанную выше процедуру массового автоматизированного оценивания состоятельности процессов принято обозначать термином — оценка «комплексного критерия состоятельности процессов» [10]. Содержательно и структурно эта процедура может быть реализована с помощью одного из вариантов класса алгоритмов ИИ, применяемых в решении подобных задач, например — метода решающих деревьев или более сложных его модификаций [31; 33; 35].

3. Кластерная структура описания процессов с учетом мультиколлинеарности факторов

Выше были кратко проанализированы большинство аспектов объективной применимости *линейной регрессионной модели* для «элементарных» процессов с точки зрения состоятельности ее параметров для описания зависимости в паре индикатор — фактор [8]. Но для описания и управления одновременно всеми «элементарными» процессами в ЦП необходима их обобщенная математическая модель. При этом важно, что ключевым критерием оптимальности такой модели является ее эффективность, т. е., по сути, количественное обоснование того, что выбор ее конкретного вида для конкретных случаев создания ЦП регионов является наилучшим вариантом аппроксимации исходных данных в классе всех потенциально применимых линейных моделей⁴. Именно в этом контексте принципиальными моментами являются выявление наличия и учет потенциальной мультиколлинеарности в *группах факторов*, выбранных и оцененных в качестве регрессоров для *одного общего индикатора*. Из этого следует, что всю совокупность уже проанализированных «элементарных» процессов необходимо структурировать в виде *кластеров*, т. е. групп парных линейных зависимостей, в которых рассматривается один и тот же индикатор, но разные факторы. Фактически это означает, что в рамках каждого кластера имеется многокомпонентная (множественная) линейная регрессия, в которой либо все регрессоры (факторы) объективно независимы, либо присутствует коллинеарность факторов в кластере.

В первом случае оценки *несмещенности, состоятельности и эффективности* множественной регрессии не требуют дополнительного, корректирующего анализа и строятся аддитивно, непосредственно с помощью соответствующих ранее полученных оценок этих характеристик для пар индикатор — фактор [28]. Во втором случае подобная корректировка необходима, чтобы исключить неинформативный шум, т. е. дополнительную погрешность, связанную с корреляцией факторов в отдельных парах и, таким образом, обосновать несмещенность и эффективность соответствующей множественной линейной зависимости. Поэтому первым шагом анализа факторов в каждом кластере является их проверка на само наличие мультиколлинеарности. Эта процедура алгоритмически должна быть основана на сравнении рассчитанных и пороговых величин $R(X_k, X_l)$ — линейных коэффициентов корреляции всех возможных пар факторов (X_k, X_l) [8]. В результате все кластеры классифицируются на две категории: первая включает группы «элементарных» процессов, где все факторы независимы, а вторая — кластеры, в которых хотя бы одна пара факторов имеет статистически значимую корреляцию. Именно вторая совокупность кластеров является предметом дальнейшего анализа и формализации соответствующей корректировки.

⁴ В математической статистике такая модель в классе всех линейных моделей обозначается термином аппроксимации с «наилучшей линейной несмещенной оценкой» или модели BLUE — Best Linear Unbiased Estimate [30].

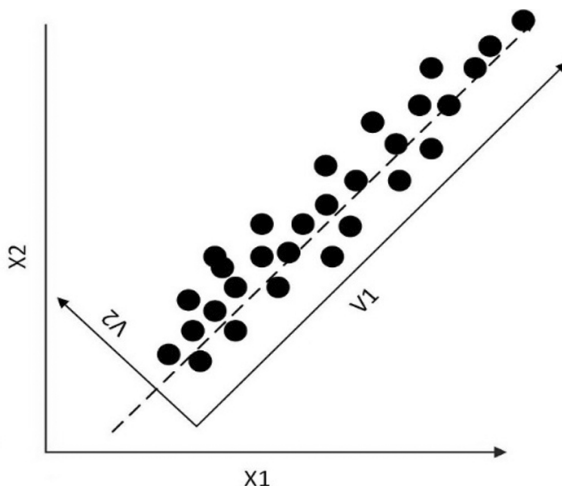


Рис. 1. Модельный пример реализаций функции $F(X_1, X_2)$ (круглые маркеры) в зависимости от коррелирующих факторов X_1 и X_2 в исходной (X_1, X_2) и модифицированной (V_1, V_2) системах координат

Fig. 1. A model example of the implementations of the function $F(X_1, X_2)$ (round markers) depending on the correlating factors X_1 and X_2 in the original (X_1, X_2) and modified (V_1, V_2) coordinate systems

Источник: составлено авторами.

Корректировки, связанные с учетом мультиколлинеарности факторов, широко используются в различных алгоритмах ИИ, преимущественно для оптимального сокращения размерности факторного пространства с минимальными потерями его информативности, но также и для эффективной фильтрации влияния неинформативных шумов и структурирования пространства индикаторов с привлечением элементов кластерного анализа, подобных примененным выше⁵.

Алгоритмической основой этих корректировок является широко применяемый метод главных компонент (МГК) [24] или эквивалентные варианты его обобщения — методы сингулярного разложения (МСР)⁶. Лежащая в основе МГК идея *эффективного учета* взаимной зависимости *аргументов* (факторов) для многокомпонентных функций типа $Y = F(X_1, X_2, \dots, X_r)$ может быть схематично проиллюстрирована с помощью модельного примера на рис. 1, где для двухкомпонентного случая $Y = F(X_1, X_2)$ *точками* показаны ее случайные реализации в системе координат коррелирующих аргументов (факторов) X_1 и X_2 .

Индивидуальные дисперсии (изменчивости) каждого из факторов X_1, X_2 , суммарно определяющие изменчивость $F(X_1, X_2)$, пропорциональны сумме проекций реализаций функции F (точек на рис. 1) соответственно на оси X_1 и X_2 . Нетрудно видеть, что на рис. 1 вклады индивидуальных дисперсий X_1 и X_2 (т. е. суммы проекций на оси X_1 и X_2 соответственно) в их суммарную дисперсию имеют примерно сопоставимую величину. Напротив, если ввести новую систему координат (V_1, V_2) , в которой ось V_1 «визуально» ориентирована по направлению максимальной суммарной изменчивости $F(X_1, X_2)$, то соответствующие проекции на оси V_1 и V_2 будут

⁵ Учет именно этих аспектов является ключевым требованием для решения задач, являющихся предметом настоящей работы, в то время как сокращение факторного пространства не столь критично.

⁶ В зарубежной литературе МГК соответствует термину Principal Component Analysis, или PCA, а МСР обозначается как Singular Value Decomposition, или SVD.

существенно различаться. При этом переход от исходной (X_1, X_2) к новой (V_1, V_2) системе координат осуществляется с помощью *линейного преобразования*, т. е. координаты точек реализации функции $F(V_1, V_2)$, а именно — $V_1(i)$ и $V_2(i)$, являются линейными комбинациями соответствующих координат в системе (X_1, X_2) :

$$\begin{aligned} V_1(i) &= \varphi_1 X_1(i) + \psi_1 X_2(i) \\ V_2(i) &= \varphi_2 X_1(i) + \psi_2 X_2(i) \end{aligned} \quad i=1, \dots, N_s, \quad (8)$$

где φ_k, ψ_k ($k = 1, 2$) *вычисляемые* параметры линейного преобразования, а N_s — количество точек в выборке реализаций функции $F(V_1, V_2)$ или $F(X_1, X_2)$ в примере на рис. 1. В частности, для этого модельного случая можно считать, что в новой системе координат при дальнейшем использовании функции $F(V_1, V_2)$ вкладом величин, связанных с аргументом V_2 , можно пренебречь без существенной потери информативности (суммарной дисперсии реализаций функции F). Но главное, можно показать, что $V_1(i)$ и $V_2(i)$ образуют *ортогональную систему координат*, в которой V_1 и V_2 являются *независимыми* переменными [24]. В многомерном случае вектора типа $V_p(i)$ определяются как *главные компоненты* модифицированного пространства аргументов (факторов). Индекс $p = 1, \dots, N_p$, указывает на номер главного компонента, причем $N_p \leq N_f$, где N_f — исходное количество аргументов (факторов в кластере), а N_p — соответствующее количество главных компонент. Например, для случая на рис. 1, т. е. при $N_f = 2$, можно считать, что $N_p = 1$.

В реальных задачах главные компоненты определяются на основе математически строгого алгоритма: как результат решения задачи нахождения собственных векторов и собственных значений двух типов матриц, содержащих значения реализаций всех факторов в кластере. В стандартном подходе МГК рассматривается их ковариационная матрица, а в МСР — матрица реализаций самих факторов. В сравнении с МГК подход МСР обеспечивает большую точность при больших размерностях массива факторов, но является более сложным с точки зрения алгоритмической реализации и вычислительных затрат. Поэтому, поскольку для рассматриваемых в работе региональных случаев наличие больших размеров матрицы реализации факторов в кластере не является критичным, представляется целесообразным использовать стандартный подход МГК, но с определенными модификациями, предложенными в [29; 34]. Ниже представлены основные детали такого применения МГК.

Прежде всего, используется эквивалентное (7) представление уравнения линейной регрессии для пары индикатора (y) и фактора (x) вида:

$$\frac{y - \bar{y}}{\bar{y}} = B \frac{(x - \bar{x})}{\bar{y}} + \frac{e}{\bar{y}}, \quad (9)$$

где B — коэффициент регрессии, e — остаточный член, а \bar{y} и \bar{x} — средние значения (математические ожидания) в выборке индексного представления типа (1)–(3) по исходным данным для пары (x_i, y_i) , $i = 1, \dots, N_{opt}$. В дискретных координатах выражение (9) можно представить с помощью следующих векторов Y, X, Q :

$$Y = X + Q, \quad (10)$$

где

$$Y = \{Y_i\} = \left\{ \frac{y_i - \bar{y}}{\bar{y}} \right\}; \quad X = \{X_i\} = \left\{ B \frac{(x_i - \bar{x})}{\bar{y}} \right\}; \quad Q = \{Q_i\} = \left\{ \frac{e_i}{\bar{y}} \right\} \quad i=1, \dots, N_{opt}.$$

Для кластера с группой факторов (например, в количестве M векторов X_k , $k = 1, \dots, M$) выражение (10) будет иметь вид:

$$Y = \sum_{k=1}^M Z_k; \quad Z_k = X_k + Q_k, \quad (11)$$

где X_k и Q_k — k -ые вектора реализаций нормированных факторов и погрешностей регрессий соответственно.

Далее, с учетом вида принятой нормировки (9)–(10) и наличия *нулевого среднего* для погрешностей в каждой k -ой регрессии в кластере [32], строятся две ковариационные матрицы (размерностью $N_{\min} \times N_{\min}$):

$$\begin{aligned} S_X &= \left\| S_{ij}^{(k)} \right\|, \quad \left\{ S_{ij}^{(k)} \right\} = \frac{1}{M-1} \sum_{k=1}^M Z_i^{(k)} Z_j^{(k)}, \\ Q_X &= \left\| Q_{ij}^{(k)} \right\|, \quad \left\{ Q_{ij}^{(k)} \right\} = \frac{1}{M-1} \sum_{k=1}^M Q_i^{(k)} Q_j^{(k)}, \\ & i, j = 1, \dots, N_{\min}; \quad k = 1, \dots, M, \end{aligned} \quad (12)$$

где S_X — ковариационная матрица факторов, модифицированная в соответствии с (9)–(11), Q_X — ковариационная матрица ошибок регрессий, нормированных в виде (10), а N_{\min} — *минимальная* имеющаяся размерность $N_{opt}(k)$ среди всех M векторов X_k в кластере.

На основе выражений (10)–(12) в модифицированном подходе к реализации МГК, используется разложение на собственные вектора (V_l) и собственные значения (λ_l) для ковариационной матрицы следующего вида:

$$\begin{aligned} G &= \|G^{ij}\| = S_X Q_X^{-1}, \\ G V_l &= \lambda_l V_l, \\ & l = 1, \dots, L, \end{aligned} \quad (13)$$

где L — число учитываемых ее собственных векторов и собственных значений (т. е. главных компонент), причем $L \leq N_{\min}$ и определяется на основе *сортировки* членов разложения (13) в порядке убывания λ_l и последующего применения критерия вида:

$$\frac{1}{L} \sum_{k=l+1}^L \lambda_k \leq 1.$$

Важно, что применимость этого критерия как раз и обеспечивается использованием модифицированной ковариационной матрицы в (13): вид G принимается исходя из требования, что погрешности аппроксимации X_k , в разложении (11) должна быть на уровне имеющейся погрешности регрессий, т. е. векторов Q_k . В этом случае величины λ_l интерпретируются как отношение дисперсии l -ой главной компоненты к дисперсии шума в направлении соответствующего собственного вектора V_l [29], а выражение для многокомпонентной зависимости (11) индикатора от факторов в кластере приобретает вид:

$$Y = \sum_{l=1}^L \lambda_l V_l. \quad (14)$$

Следует отметить, что в (14) все элементы векторов V_l , т. е. $V_l(i)$, $i = 1, \dots, N_{\min}$, являются линейными комбинациями вида (8) соответствующих элементов

$x_k(i)$ ($k = 1, \dots, M$) исходных, заданных в форме индексов типа (1)–(3) факторов. При этом собственные значения λ_i и параметры нормировок, используемых в (9), по определению также являются *вычисляемыми величинами*. Исходя из того, что все преобразования переменных, используемые в (9)–(14), *линейны*, то к (14) применимо обратное преобразование координат. Иными словами, математически обоснован *переход* от представления многокомпонентной зависимости индикатора от нескольких факторов в кластере в виде векторного разложения (14) к эквивалентному виду непрерывной эффективной зависимости индикатора y от факторов x_k для кластера, содержащего M факторов:

$$y = C_1 x_1 + C_2 x_2 \dots + C_M x_M. \quad (15)$$

Коэффициенты C_k , $k = 1, \dots, M$ являются *численными параметрами* линейной модели (15), однозначно *вычисляемыми* на основе разложений (9)–(14), а возможности такого вычисления обеспечиваются указанной линейной зависимостью компонентов собственных векторов V_1 и векторов X_k . При этом важно, что значения коэффициентов C_k рассчитываются на основе параметров разложения МГК (причем с модификациями, учитывающими ковариации погрешностей для исходных зависимостей в разных парах индикатор — фактор). Это позволяет в (15) учитывать влияние *мультиколлинеарности факторов* в кластере и, таким образом, обеспечить *эффективность* линейной многокомпонентной регрессии индикатора для случаев коррелирующих факторов.

Кроме того, наличие сопоставимости численных масштабов величин индикатора и всех факторов в кластере обеспечивает возможность рассматривать C_k как весовые оценки вклада отдельных x_k в формирование значения y . На этой основе могут быть установлены и опционально применены определенные пороговые критерии, позволяющие при необходимости пренебрегать вкладом отдельных факторов в значение индикатора y без существенной потери точности его оценки с помощью (15).

Таким образом, вид зависимости в выражении (15) и, главное, описанный способ количественного оценивания коэффициентов C_k следует рассматривать в качестве исходных методических условий для выбора оптимального алгоритмического подхода для математических моделей описания и управления совокупностью процессов во всех типах кластеров.

4. Экономико-математическое моделирование управления процессами

Обоснованность описания зависимостей индикатора y от нескольких факторов x_k с помощью выражений вида (15) существенно расширяет возможности построения модели управления процессами на основе эффективных методов экономического факторного анализа (ЭФА) [5]. Целесообразность построения оптимальной математической модели с применением подобных методов обусловлена следующими аргументами.

Модели ЭФА, основанные на развитии многомерных методов математической статистики, линейной алгебры и классического математического анализа, предназначены для количественного описания прежде всего *содержательной связи* между индикаторами и отдельными факторами [25], т. е. для нахождения способов *управления изменениями индикаторов* процессов с помощью математически детализированных зависимостей этих изменений *от изменений отдельных факторов*. Это отличает методический подход ЭФА, например, от моделей на основе ИНС, нацеленных исключительно на максимизацию объясненной дисперсии индикатора с помощью оптимизации совокупного, содержательно не детализируемого набора параметров функций, формально связывающих индикатор с факторами, а также факторы друг с другом [4; 33; 35]. С другой стороны, диапазон практического

применения современных моделей ЭФА зачастую ограничен методическими трудностями выбора структуры, а также способов вычисления параметров так называемой «производственной функции» процесса, т. е. определенной формы математического отображения пространства значений индикатора процесса (например, объемов выпуска продукции) на пространство заданных значений набора связанных с этим процессом факторов (например, ресурсов разного типа) [5; 18]. Однако применительно к решению задачи построения модели управления процессами в рамках создания ЦП эта проблема отсутствует, поскольку в этом случае в качестве «производственной функции» может очевидно рассматриваться многокомпонентная линейная зависимость индикатора от факторов с *рассчитываемыми коэффициентами* вида (15).

Для построения модели управления в качестве оптимального подхода ЭФА целесообразно использовать *метод конечных приращений* (МКП или метод Лагранжа), позволяющий получать в отдельных прикладных случаях точные формулы для расчета влияния изменений группы факторов на изменение общего индикатора в кластере процессов, причем при конечных (т. е. не обязательно малых) величинах этих изменений [5]. Именно к таким случаям относится вариант использования в МКП «производственной функции» в виде (15).

Как известно, метод Лагранжа строится на основе использования положений двух теорем классического математического анализа, а именно — теоремах о среднем значении дифференциального и интегрального исчисления, которые в качестве условий применимости требуют для рассматриваемых функций зависимости индикатора от факторов лишь дифференцируемости и интегрируемости в пределах изменений соответствующих показателей, что очевидно обеспечивается характером используемого выражения (15) — линейной зависимости с известными коэффициентами. Применительно к кластеру процессов в общей формулировке алгоритма МКП изменение индикатора Δy в зависимости от приращений факторов Δx_k может быть представлено в виде

$$\Delta y = \sum_k^M f'_{x_k}(\tilde{x}_k) \Delta x_k, \quad (16)$$

$$f'_{x_k}(\tilde{x}_k) = \frac{\partial y(\tilde{x}_k)}{\partial x_k},$$

где частные производные вычисляются для фиксированных значений факторов \tilde{x}_k , в которых (исходя из положений теоремы Лагранжа) достигается точное разложение приращения общего индикатора y [5].

Тогда на основе (16), с учетом выражения (15) для кластеров с изначально коррелирующими факторами, а также аддитивности вкладов приращений Δx_k в Δy для кластеров с независимыми факторами [10] может быть сформирована следующая универсальная модель управления процессами:

$$\Delta y = \sum_k^M W_k \Delta x_k, \quad (17)$$

где $W_k = C_k$ — коэффициенты разложения (15) для случая изначально коррелирующих факторов; и $W_k = B_k$ — коэффициенты регрессии линейных зависимостей от x_k для кластеров с независимыми факторами соответственно.

Следует отметить, что для любых категорий кластеров модель (17) эффективна в статистическом смысле [32], обеспечивает получение оценок Δy при любых конечных приращениях факторов Δx_k [5], а также приемлемую, количественно обоснованную точность для построения краткосрочных прогнозов вне поля корреляции

показателей, используемых при ее построении. Единственной методической сложностью в этом контексте является выбор применимых (т. е. эффективных для каждого конкретного случая) алгоритмов построения доверительных интервалов прогнозной оценки математического ожидания y [10]. Как известно, для этого существует целый ряд соответствующих алгоритмических подходов [8], однако оценивание деталей эффективности их применения в различных случаях не представляется существенным аспектом в настоящей работе.

Также важно, что в выражениях (16)–(17) все используемые переменные формализованы в индексном представлении. С одной стороны, это обеспечивает универсальность и сопоставимость получаемых с ее помощью оценок применительно к региональным системам различного масштаба и уровня социально-экономических показателей (в абсолютных единицах). С другой стороны, моделируемые с помощью (17) изменения индикаторов являются, по сути, относительными характеристиками динамики процессов, что должно соответствующим образом учитываться при использовании (17) в рамках модели обеспечения принятия управленческих решений.

5. Сценарные модели стратегического управления процессами на основе принципов конвергенции естественного и искусственного интеллекта

Новый уровень обоснованности принятия управленческих решений должен обеспечиваться многоаспектными моделями выполнения всех потенциальных сценариев управления процессами. При этом, в соответствии с принципами концепции конвергенции «человек — искусственный интеллект», эти модели, с одной стороны, должны быть снабжены адаптивным интерфейсом, учитывающим объективные административные требования подготовки решений стратегического управления, а с другой — обеспечивать все необходимые *количественные характеристики* социально-экономических эффектов при анализе и выборе того или иного сценария управления.

Таким образом, представляется целесообразным структурно формировать подобные модели на основе трех *взаимосвязанных* функциональных блоков: факторно-аналитического, логико-математического и интерфейсного соответственно.

В частности, факторно-аналитическая часть строится на основе кластерной математической модели, состоящей из процедуры структурирования в кластеры зависимостей одного и того же индикатора от нескольких факторов; процедуры идентификации типов кластеров с точки зрения независимости рассматриваемых в кластере факторов; а также собственно самой универсальной математической модели вида (17).

В свою очередь, логико-математический блок моделей состоит из аналитической процедуры структурирования кластеров с точки зрения *привязки их индикаторов к конкретным параметрам стратегических целей* развития регионов; процедуры формирования на этой основе потенциальных сценариев управления процессами в кластерах; а также процедуры выбора (на основе автоматизированного целевого анализа потенциальных сценариев) оптимального комплекса управленческих решений в зависимости от объективного наличия ресурсов для их реализации.

Наконец, интерфейсный блок модели должен включать в себя два функциональных элемента: процедуру преобразования количественных параметров оптимального сценария из индексного представления в показатели, выраженные в естественных (абсолютных) единицах, а также процедуру представления оптимального сценария в соответствии с формальными требованиями управленческих регламентов, относящихся к сфере институциональной ответственности региональных администраций.

С методической точки зрения реализация факторно-аналитической части моделей предполагает использование автоматизированных процедур кластерного анализа

исходных зависимостей индикатор — фактор, связанных с применением класса алгоритмов ИИ, известных как классификация с помощью фиксированных правил [33; 35], и преимущественно основанных на использовании математико-статистических оценок [10; 26]. В свою очередь, методическое обеспечение логического и отчасти интерфейсного блоков должно основываться на одном из вариантов соответствующего класса алгоритмов ИИ, известного как метод решающих деревьев, или на основе его различных детализирующих или обобщающих модификаций [31].

Следует отметить общую методическую особенность реализации всех трех функциональных блоков подобных моделей, а именно — широкое использование методов, так или иначе связанных с определенными классами алгоритмов ИИ.

Результаты и выводы

Основные выводы проведенного в работе анализа можно кратко сформулировать следующим образом.

1. Ключевым условием достижения стратегических целей развития России является разработка и реализация цифровых платформ управления сложными социально-экономическими системами, основанными на их цифровой трансформации, прежде всего, на региональном уровне. В свою очередь, цифровая трансформация систем обусловлена необходимостью создания их цифровых портретов (ЦП), что является многоаспектной и многоэтапной задачей, поэтому предметом настоящей работы стал анализ различных информационно-технологических и математических подходов для различных этапов ее решения, оптимальных, прежде всего, с точки зрения наличия исходных статистических данных, характеризующих социально-экономические процессы в регионах.

2. Обширность исходного массива региональной статистики требует использования методологии интеллектуального анализа больших данных (Data Mining), сочетающей экспертно-теоретические и математические аналитические методы, прежде всего, алгоритмы искусственного интеллекта (ИИ). Однако данные региональной статистики обладают структурной спецификой, связанной с большим количеством различного типа показателей, характеризующих региональную систему в целом, при достаточно малом объеме актуальных данных по каждому отдельному показателю, что исключает применение алгоритмов ИИ на основе технологий ИНС.

3. Обоснована оптимальность использования алгоритмов ИИ в сочетании с линейными моделями описания региональных процессов, основанных на классических методах математической статистики и индексном представлении анализируемых показателей. При этом моделирование процессов использует кластерный подход, где каждый кластер связан с множественной регрессионной зависимостью одного индикатора от нескольких факторов.

4. Выполнен анализ способов учета корреляции факторов в кластере с помощью метода главных компонент (МГК) с использованием модифицированной ковариационной матрицы факторов и погрешностей парных регрессий индикатор — фактор. В результате построена эффективная зависимость индикатора от факторов кластера в виде линейного разложения с фиксированными коэффициентами, учитывающими влияние мультиколлинеарности.

5. Предложена универсальная экономико-математическая модель управления процессами, совместно использующая один из подходов экономического факторного анализа (метод конечных приращений) и зависимости индикатора от факторов в кластерах в виде линейного разложения с фиксированными коэффициентами.

6. Описана методология построения сценарных моделей стратегического управления на основе принципов конвергенции естественного и искусственного интеллекта. Предложен состав необходимых процедур, описаны назначения и характеристики

для каждого из блоков моделей, а также методы их реализации на основе алгоритмов ИИ.

7. В совокупности представленные результаты демонстрируют реальные возможности для разработки и практической реализации информационно-технологической платформы ЦП социально-экономических систем регионов России как одного из необходимых элементов *стратегического управления процессами* достижения национальных целей их развития.

Заключение

В условиях стремительного развития цифровых технологий, ориентированных на решение управленческих задач, стратегически важным является *формирование результативных отношений* между субъектами соответствующего направления научных исследований и государственными органами, поскольку вклад подобных научных разработок в сферу цифровой трансформации государственных систем управления становится решающим фактором социального и экономического прогресса. В настоящей работе аргументируется необходимость развития систем управления за счет использования новых цифровых методов анализа социально-экономических процессов, в том числе методов, основанных на алгоритмах искусственного интеллекта.

Однако в современной практике по-прежнему присутствует заметный разрыв между потенциальными возможностями повышения эффективности систем регионального управления за счет внедрения цифровых технологий и объективно достигнутым уровнем практического использования этих технологий. Именно по этой причине при обосновании оптимальности использования рассмотренных в работе методов и моделей основной акцент сделан на возможности их практического применения в реальных системах. Этот акцент указывает на то, что приоритетными факторами, определяющими необходимость модернизации существующих систем, должны стать новые возможности цифрового управления, созданные в результате соответствующих научных исследований, а не потребности частичного усовершенствования существующих управленческих процедур государственного администрирования.

Литература

1. Адамов В. Е. Факторный индексный анализ (методология и проблемы). М. : Статистика, 1977. 200 с.
2. Акаев А. А., Ичкитидзе Ю. Р., Петряков А. А., Сарыгулов А. И. Цифровая трансформация экономики: эмпирические факты и математические модели. СПб. : Издательско-полиграфическая ассоциация высших учебных заведений, 2020. 336 с.
3. Акаев А. А., Садовничий В. А. Человеческий фактор как определяющий производительность труда в эпоху цифровой экономики // Проблемы прогнозирования. 2021. № 1 (184). С. 45–58. DOI 10.47711/0868-6351-184-45-58. EDN WFYDGO
4. Белоглазов Д. А. Особенности нейросетевых решений, достоинства и недостатки, перспективы применения // Известия Южного федерального университета. Технические науки. 2008. Т. 84, № 7. С. 105–110. EDN KAPCWZ
5. Блюмин С. Л., Суханов В. Ф., Чеботарёв С. В. Экономический факторный анализ. Липецк : Изд-во Липецкого эколого-гуманитарного института (ЛЭГИ). 2004. 148 с.
6. Виноградова Н. М. Теория индексов. М. : Гостехиздат, 1930. 200 с.
7. Дюк В. А., Флегонтов А. В., Фомина И. К. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях // Известия РГПУ им. А. И. Герцена. 2011. № 138. С. 77–83. EDN NDNWEJ
8. Елисеева И. И., Князевский В. С., Ниворожкина Л. И., Морозова З. А. Теория статистики с основами теории вероятности. М. : Финансы и статистика. 2002. 400 с.

9. Ефанов В. А., Чаадаев В. К., Шляхов А. С. Стратегирование цифровой трансформации промышленного предприятия (на примере ФГУП «Российская телевизионная и радиовещательная сеть») // Экономика промышленности. 2023. Т. 16, № 1. С. 95–104. DOI 10.17073/2072-1633-2023-1-95-104. EDN ADVHOJ
10. Журавлев Д. М. Стратегирование цифровой трансформации сложных социально-экономических систем / под науч. ред. В. Л. Квинта. СПб. : ИПЦ СЗИУ РАНХиГС, 2024. (Серия «Библиотека стратега»). 352 с.
11. Журавлев Д. М., Троценко А. Н., Чаадаев В. К. Методология и инструментарий стратегирования социально-экономического развития региона // Экономика промышленности. 2022. Т. 15, № 2. С. 131–142. DOI 10.17073/2072-1633-2022-2-131-142. EDN IAXLHE
12. Журавлев Ю. И., Рязанов В. В., Сенько О. В. Распознавание. Математические методы. Программная система. Практические применения. М. : Фазис. 2006. 176 с.
13. Журавлев Ю. И., Флеров Ю. А., Вялый М. Н. Основы высшей алгебры и теории кодирования. М. : Изд-во ФУПМ МФТИ, 2019. 308 с.
14. Замков О. О., Толстопятенко А. В., Черемных И. Н. Математические методы в экономике. М. : Дело и сервис, 1997. 368 с.
15. Ивченко Г. И., Медведев Ю. И. Математическая статистика : учебник. М. : КД «ЛИБРОКОМ», 2014. 352 с.
16. Исмагилов И. И., Кадочникова Е. И., Костромин А. В. Эконометрика. Казань : Изд-во Казанского университета, 2014. 235 с.
17. Капелюшников Р. И. Искусственный интеллект и проблема сингулярности в экономике // Препринт WP3. 2025.01. Серия WP3 «Проблемы рынка труда». М. : Изд. дом Высшей школы экономики, 2025. 67 с.
18. Клейнер Г. Б. Производственные функции. Теория, методы, применение. М. : Финансы и статистика, 1986. 238 с.
19. Кремер Н. Ш., Путько Б. А. Эконометрика : учебник для вузов. М. : ЮНИТИ-ДАНА, 2004. 311 с.
20. Лаптева Е. А., Навдаева С. Н., Ирхина Л. Н. Статистика: индексный метод анализа : учеб. пособие. Н. Новгород : Изд-во Нижегородского ГАТУ, 2022. 164 с.
21. Меликян А. А. Применение индексного метода в исследовании региональной цифровой дифференциации // Инновации и инвестиции. 2025. № 3. С. 406–409. EDN GINLLH
22. Некипелов А. Д. Об экономической стратегии и экономической политике России в современных условиях // Научные труды Вольного экономического общества России. 2021. Т. 230, № 4. С. 76–89. DOI 10.38197/2072-2060-2021-230-4-76-89. EDN MHLEYL
23. Некипелов А. Д. От нейтрализации внешних шоков к устойчивому долгосрочному развитию // Научные труды Вольного экономического общества России. 2024. Т. 248, № 4. С. 130–142. DOI 10.38197/2072-2060-2024-248-4-130-142. EDN IMGYWO
24. Обухов А. М. О статистических ортогональных разложениях эмпирических функций // Известия АН СССР. Сер. Геофизика. 1960. № 3. С. 432–439.
25. Овсянников Г. Н. Факторный анализ в доступном изложении: Изучение многопараметрических систем и процессов. М. : КД «ЛИБРОКОМ», 2025. 176 с.
26. Орлов А. И. Искусственный интеллект: статистические методы анализа данных : учебник. М. : Ай Пи Ар Медиа, 2022. 843 с.
27. Регионы России: социально-экономические показатели. М. : Росстат, 2020. 1242 с.
28. Трофимова Е. А., Кисляк Н. В., Гилев Д. В. Теория вероятностей и математическая статистика : учеб. пособие. Екатеринбург : Изд. Уральского университета, 2018. 160 с.
29. Успенский А. Б., Романов С. В., Троценко А. Н. Применение метода главных компонент для анализа ИК-спектров высокого разрешения, измеренных со спутников // Исследования Земли из космоса. 2003. № 3. С. 26–33. EDN OOCSSX
30. Aggarwal Charu C. Neural Networks and Deep Learning: A Textbook, Second Edition, 2023. Springer Cham, 2024. 529 p.
31. Breiman L. Bagging Predictors // Machine Learning. 1996. № 24. P. 123–140.
32. Cook R. D., Weisberg S. Residuals and Influence in Regression. New York: Chapman and Hall, 1982. 230 p.
33. Goodfellow I., Bengio Y., Courville A. Deep Learning (Adaptive Computation and Machine Learning series). The MIT Press, 2016. 800 p.
34. Kao Yi-Hao, Van Roy B. Directed Principal Component Analysis // Operations Research. 2014. Vol. 62, N 4. P. 957–972.
35. McAuley J. Personalized Machine Learning. Cambridge University Press, 2022. 326 p.

36. *Stephens-Davidowitz S.* Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. Dey Street Books, 2018. 352 p.
37. *Zaki, M. J., Wagner Meira Jr.* Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Second Edition. Publisher: Cambridge University Press. 2020. 766 p.

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Об авторе:

Троценко Анатолий Николаевич, доктор физико-математических наук, член Ученого совета Научно-исследовательского института Социальных Систем при МГУ имени М. В. Ломоносова (Москва, Российская Федерация); info@niiss.ru

References

1. Adamov V. E. Factor index analysis (methodology and problems). Moscow: Statistika Publishing House, 1977. 200 p. (In Russ.).
2. Akaev A. A., Ichkitidze Yu. R., Petryakov A. A., Sarygulov A. I. Digital transformation of the economy: empirical facts and mathematical models. St. Petersburg: Publishing and Printing Association of Higher Educational Institutions, 2020. 336 p. (In Russ.).
3. Akaev A. A., Sadovnichy V. A. The human factor as a determinant of labor productivity in the era of the digital economy // Problems of Forecasting [Problemy prognozirovaniya]. 2021. N 1 (184). P. 45–58. DOI 10.47711/0868-6351-184-45-58. EDN WFDGO. (In Russ.).
4. Beloglazov D. A. Features of neural network solutions, advantages and disadvantages, application prospects // Bulletin of the Southern Federal University. Technical sciences [Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki]. 2008. Vol. 84, N 7. P. 105–110. EDN KAPCWZ. (In Russ.).
5. Blumin S. L., Sukhanov V. F., Chebotarev S. V. Economic factor analysis. Lipetsk: Publishing house of the Lipetsk Ecological and Humanitarian Institute. 2004. 148 p. (In Russ.).
6. Vinogradova N. M. Theory of indexes. Moscow: Gostekhzdat, 1930. 200 p. (In Russ.).
7. Duke V. A., Flegontov A. V., Fomina I. K. Application of data mining technologies in the natural sciences, engineering and humanitarian fields // Bulletin of the Herzen State Pedagogical Univ. of Russia [Izvestiya RGPU im. A. I. Gertsena]. 2011. N 138. P. 77–83. EDN NDNWEJ. (In Russ.).
8. Eliseeva I. I., Kryazevsky V. S., Nivorozhkina L. I., Morozova Z. A. Theory of statistics with the basics of probability theory. Moscow: Finance and Statistics, 2002. 400 p. (In Russ.).
9. Efanov V. A., Chaadaev V. K., Shlyakhov A. S. Strategizing the digital transformation of an industrial enterprise (on the example of the Russian Television and Radio Broadcasting Network Federal State Unitary Enterprise) // Industrial Economics [Ekonomika promyshlennosti]. 2023. Vol. 16, N 1. P. 95–104. DOI 10.17073/2072-1633-2023-1-95-104. EDN ADVHOJ. (In Russ.).
10. Zhuravlev D. M. Strategizing the Digital Transformation of Complex Socioeconomic Systems / edited by V. L. Kvint. St. Petersburg: NWIM of RANEPa, 2024. (Series “Strategist’s Library”). 352 p. (In Russ.).
11. Zhuravlev D. M., Trotsenko A. N., Chaadaev V. K. Methodology and Tools for Strategizing the Socioeconomic Development of a Region // Industrial Economics [Ekonomika promyshlennosti]. 2022. Vol. 15, N 2. P. 131–142. DOI 10.17073/2072-1633-2022-2-131-142. EDN IAXLHE. (In Russ.).
12. Zhuravlev Yu. I., Ryazanov V. V., Senko O. V. Recognition. Mathematical Methods. Software System. Practical Applications. Moscow: Phasis Publishing House. 2006. 176 p. (In Russ.).
13. Zhuravlev Yu. I., Flerov Yu. A., Vyal'yi M. N. Fundamentals of Higher Algebra and Coding Theory. Moscow: Publishing House of the Faculty of Management and Applied Mathematics of the Moscow Institute of Physics and Technology, 2019. 308 p. (In Russ.).
14. Zamkov O. O., Tolstopiatenko A. V., Cheremnykh I. N. Mathematical Methods in Economics. Moscow: Delo i Servis Publishing House, 1997. 368 p. (In Russ.).
15. Ivchenko G. I., Medvedev Yu. I. Mathematical Statistics: Textbook. Moscow: LIBROKOM Publishing House, 2014. 352 p. (In Russ.).
16. Ismagilov I. I., Kadochnikova E. I., Kostromin A. V. Econometrics. Kazan: Kazan University Publishing House, 2014. 235 p. (In Russ.).
17. Kapelyushnikov R. I. Artificial Intelligence and the Problem of Singularity in Economics // Preprint WP3. 2025.01. Series WP3 “Problems of the Labor Market”. Moscow: Publishing House of the Higher School of Economics, 2025. 67 p. (In Russ.).

18. Kleiner G. B. Production Functions. Theory, Methods, Application. Moscow: Finance and Statistics Publishing House, 1986. 238 p. (In Russ.).
19. Kremer N. Sh., Putko B. A. Econometrics: Textbook for Universities. Moscow: UNITY-DANA Publishing House, 2004. 311 p. (In Russ.).
20. Lapteva E. A., Navdaeva S. N., Irkhina L. N. Statistics: Index Method of Analysis: Tutorial. Nizhny Novgorod: Publishing house of Nizhny Novgorod State Technical University, 2022. 164 p. (In Russ.).
21. Melikyan A. A. Application of the index method in the study of regional digital differentiation // Innovations and Investments [Innovatsii i investitsii]. 2025. N 3. P. 406–409. EDN GINLLH. (In Russ.).
22. Nekipelov A. D. On the economic strategy and economic policy of Russia in modern conditions // Scientific works of the VEO of Russia [Nauchnye trudy VEO Rossii]. 2021. Vol. 230, N 4. P. 76–89. DOI 10.38197/2072-2060-2021-230-4-76-89. EDN MHLEYL. (In Russ.).
23. Nekipelov A. D. From neutralization of external shocks to sustainable long-term development // Scientific works of the VEO of Russia [Nauchnye trudy VEO Rossii]. 2024. Vol. 248, N 4. P. 130–142. DOI 10.38197/2072-2060-2024-248-4-130-142. EDN IMGWVO. (In Russ.).
24. Obukhov A. M. On statistical orthogonal expansions of empirical functions // Bulletin of the USSR Academy of Sciences. Series: Geophysics [Izvestiya AN SSSR. Ser. Geofizika]. 1960. N 3. P. 432–439. (In Russ.).
25. Ovsyannikov G. N. Factor analysis in an accessible presentation: Study of multiparameter systems and processes. Moscow: LIBROKOM Publishing House, 2025. 176 p. (In Russ.).
26. Orlov A. I. Artificial Intelligence: Statistical Methods of Data Analysis: Textbook. Moscow: IPR Media Publishing House, 2022. 843 p. (In Russ.).
27. Regions of Russia: Socioeconomic Indicators. Moscow: Rosstat, 2020. 1242 p. (In Russ.).
28. Trofimova E. A., Kislyak N. V., Gilev D. V. Probability Theory and Mathematical Statistics: Textbook. Yekaterinburg: Ural University Press, 2018. 160 p. (In Russ.).
29. Uspensky A. B., Romanov S. V., Trotsenko A. N. Application of the Principal Component Analysis for the Analysis of High-Resolution IR Spectra Measured from Satellites // Research of the Earth from Space [Issledovaniya Zemli iz kosmosa]. 2003. N 3. P. 26–33. EDN OOCSSX. (In Russ.).
30. Aggarwal Charu C. Neural Networks and Deep Learning: A Textbook, Second Edition, 2023. Springer Cham, 2024. 529 p.
31. Breiman L. Bagging Predictors // Machine Learning. 1996. № 24. P. 123–140.
32. Cook R. D., Weisberg S. Residuals and Influence in Regression. New York: Chapman and Hall, 1982. 230 p.
33. Goodfellow I., Bengio Y., Courville A. Deep Learning (Adaptive Computation and Machine Learning series). The MIT Press, 2016. 800 p.
34. Kao Yi-Hao, Van Roy B. Directed Principal Component Analysis // Operations Research. 2014. Vol. 62, N 4. P. 957–972.
35. McAuley J. Personalized Machine Learning. Cambridge University Press, 2022. 326 p.
36. Stephens-Davidowitz S. Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. Dey Street Books, 2018. 352 p.
37. Zaki, M. J., Wagner Meira Jr. Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Second Edition. Publisher: Cambridge University Press. 2020. 766 p.

Conflict of interests

The author declares no relevant conflict of interests.

About the author:

Anatoly N. Trotsenko, Doctor of Physics and Mathematics, Science Councilman in Research Institute for Social Systems at Lomonosov Moscow State University (Moscow, Russian Federation); info@niiss.ru

Поступила в редакцию: 03.12.2025
 Поступила после рецензирования: 12.12.2025
 Принята к публикации: 12.01.2026

The article was submitted: 03.12.2025
 Approved after reviewing: 12.12.2025
 Accepted for publication: 12.01.2026

© Троценко А. Н., 2026