

УДК 631.1,62-5

5.2.2 Математические, статистические и инструментальные методы экономики (физико-математические науки, экономические науки)

ИССЛЕДОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ И МЕТОДОВ ПЕРСОНАЛИЗИРОВАННОГО ОЗДОРОВЛЕНИЯ И РАЗРАБОТКА МАТЕМАТИЧЕСКОГО АППАРАТА ДЛЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПЕРСОНАЛИЗИРОВАННОГО ОЗДОРОВЛЕНИЯ ДЕТЕЙ С ИСПОЛЬЗОВАНИЕМ МОДУЛЯ СТАТИСТИЧЕСКОЙ ОТЧЕТНОСТИ

Лаптев Сергей Владимирович
к.ф.-м.н., доцент
Кубанский государственный аграрный университет, Россия, 350044, Краснодар, Калинина 13

Пигарев Марк Романович
Студент
Кубанский государственный аграрный университет, Россия, 350044, Краснодар, Калинина 13

Основной целью проводимого исследования является разработка математической модели для информационной системы персонализированного оздоровления детей, интегрирующей модули статистического анализа медико-педагогических данных и автоматизированной генерации отчётов, позволяющей повысить обоснованность назначения оздоровительных программ и снизить трудоёмкость ведения учётно-отчётной документации. В ходе исследования последовательно проанализированы существующие подходы к информатизации детского оздоровления, формализована задача персонализированного подбора оздоровительных программ, обоснован выбор статистических методов и алгоритмов рекомендаций

КЛЮЧЕВЫЕ СЛОВА:
ПЕРСОНАЛИЗИРОВАННОЕ ОЗДОРОВЛЕНИЕ, СТАТИСТИЧЕСКИЙ АНАЛИЗ, РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА, МАТЕМАТИЧЕСКАЯ МОДЕЛЬ, ФУНКЦИЯ СООТВЕТСТВИЯ, МЕТРИКИ КАЧЕСТВА

<http://dx.doi.org/10.21515/1990-4665-221-014>

UDC 631.1,62-5

5.2.2 Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

INVESTIGATION OF MATHEMATICAL MODELS AND METHODS FOR PERSONALIZED HEALTH IMPROVEMENT AND DEVELOPMENT OF THE MATHEMATICAL FRAMEWORK FOR A PERSONALIZED CHILD HEALTH IMPROVEMENT INFORMATION SYSTEM INCORPORATING A STATISTICAL REPORTING MODULE

Laptev Sergey Vladimirovich
Cand.Phys.-Math.Sci., Associate Professor
Kuban State Agrarian University, Russia, 350044, Krasnodar, Kalinina 13

Pigarev Mark Romanovich
Student
Kuban State Agrarian University, Russia, 350044, Krasnodar, Kalinina 13

The primary objective of this study is to develop a mathematical model for an information system designed for the personalized health improvement of children. This system integrates modules for the statistical analysis of medical and pedagogical data and for automated report generation, thereby enhancing the rationale behind the prescription of health improvement programs and reducing the workload associated with record-keeping and reporting. The study involved a systematic analysis of existing approaches to the informatization of child health improvement, the formalization of the task of personalized program selection, and the justification of the statistical methods and recommendation algorithms employed

Keywords: PERSONALIZED WELLNESS, STATISTICAL ANALYSIS, RECOMMENDER SYSTEM, MATHEMATICAL MODEL, MATCHING FUNCTION, QUALITY METRICS

В Российской Федерации вопросы сохранения и укрепления здоровья детского населения приобретают особую актуальность в условиях неблагоприятной динамики показателей физического развития и хронизации заболеваний среди детей школьного возраста. По данным Министерства здравоохранения РФ, доля детей с хроническими заболеваниями за последние десять лет увеличилась с 16 % до 22 %, а число детей, относящихся к первой и второй группам здоровья, неуклонно сокращается. Санаторно-оздоровительные учреждения остаются одним из ключевых звеньев системы детского здравоохранения: ежегодно в российских детских санаторно-оздоровительных комплексах проходят лечение и реабилитацию сотни тысяч детей, при этом на каждый регион приходится от пяти до двадцати учреждений различного профиля и вместимости.

В сфере медицинской информатики активно развиваются подходы к персонализации лечебно-оздоровительных мероприятий на основе методов статистического анализа и интеллектуальной обработки данных. Доказано, что учёт индивидуальных особенностей ребёнка – возраста, группы здоровья, анамнеза, результатов функциональных проб – при назначении оздоровительных процедур повышает их эффективность на 20–30 % по сравнению с унифицированными программами. Однако существующие информационные системы медицинского назначения (Медиалог, Инфоклиника, ЕСМИА и другие) ориентированы преимущественно на амбулаторно-поликлиническое и стационарное звено, обладают избыточной функциональностью и высокой стоимостью внедрения, что делает их малоприспособленными для детских санаторно-оздоровительных комплексов со специфической организацией работы.

Целью исследования является разработка математической модели для информационной системы персонализированного оздоровления детей,

интегрирующей модули статистического анализа медико-педагогических данных.

Объект исследования – методы и средства информационной поддержки персонализированного оздоровления детей в учреждениях здравоохранения и образования.

Предмет исследования – применение методов статистического анализа данных о состоянии здоровья детей и алгоритмов персонализированных рекомендаций для повышения эффективности оздоровительных мероприятий.

Для достижения цели исследования должны быть решены следующие основные задачи:

- проанализировать существующие подходы к информатизации детского оздоровления;
- разработать математические модели персонализированного подбора оздоровительных программ на основе медико-педагогических данных о состоянии здоровья детей;
- обосновать выбор статистических методов для анализа динамики показателей здоровья и оценки эффективности оздоровительных мероприятий

При проведении исследования использован комплекс методов, сочетающий теоретические и прикладные подходы:

- математическое моделирование – формализация задачи персонализированного подбора оздоровительных программ, построение функций соответствия профиля ребёнка и параметров программ;
- статистические методы – описательная статистика, корреляционный и регрессионный анализ для выявления значимых факторов оздоровления и оценки динамики показателей здоровья;

– методы рекомендательных систем – взвешенная оценка сходства профилей, ранжирование программ по степени соответствия индивидуальным показателям;

– экспериментальная оценка – сравнение качества рекомендаций персонализированной системы с унифицированным назначением, измерение временных характеристик формирования отчётов.

Выбранная совокупность методик обеспечивает одновременно научную строгость исследования.

Основные множества и обозначения

Для формализации задачи персонализированного оздоровления введём систему множеств и обозначений, отражающих ключевые сущности предметной области детского санаторно-оздоровительного комплекса.

Пусть $D = \{d_1, d_2, \dots, d_n\}$ – множество детей, проходящих оздоровление в ДСОК, где n – общее число детей в текущей смене. Каждый ребёнок $d_i \in D$ характеризуется профилем здоровья, представляющим собой вектор признаков в многомерном пространстве. Обозначим множество признаков здоровья как $F = \{f_1, f_2, \dots, f_m\}$, где m – число признаков, учитываемых при подборе программы. Тогда профиль здоровья i -го ребёнка описывается вектором $p_i = (p_{i1}, p_{i2}, \dots, p_{im})$, где p_{ij} – значение j -го признака для i -го ребёнка. Множество всех профилей образует матрицу P размерности $n \times m$.

Множество признаков F структурно разделено на четыре подмножества, соответствующих различным аспектам состояния здоровья ребёнка. Номинальные, или категориальные, признаки объединены в подмножество $F^{\text{nom}} = \{f_1, \dots, f_k\}$. Сюда относятся группа здоровья (I–V), основной диагноз по МКБ-10, пол, наличие аллергических реакций и тип нарушения осанки.

Порядковые признаки образуют подмножество $F^{\text{ord}} = \{f_{k+1}, \dots, f_l\}$; в него входят степень нарушения зрения, уровень физического развития (ниже среднего, средний, выше среднего) и степень адаптации к климатическим условиям.

Количественные непрерывные признаки собраны в подмножестве $F^{\text{cont}} = \{f_{l+1}, \dots, f_r\}$. Это возраст (в годах), рост (в см), масса тела (в кг), индекс массы тела, результаты пробы Руфье и показатели спирометрии.

Наконец, бинарные признаки составляют подмножество $F^{\text{bin}} = \{f_{r+1}, \dots, f_m\}$. К ним отнесены наличие или отсутствие хронического заболевания, наличие или отсутствие противопоказаний к конкретным процедурам и статус вакцинации. Таким образом, $F = F^{\text{nom}} \cup F^{\text{ord}} \cup F^{\text{cont}} \cup F^{\text{bin}}$, и подмножества попарно не пересекаются.

Пусть $C = \{c_1, c_2, \dots, c_s\}$ – множество оздоровительных процедур, доступных в конкретном ДСОК, где s – число процедур. Каждая процедура $c_j \in C$ описывается вектором параметров $q_j = (q_{j1}, \dots, q_{jt})$, включающим: целевые группы здоровья, минимальный и максимальный возраст назначения, профиль заболеваний (показания), перечень противопоказаний, длительность одной процедуры, максимально допустимое количество процедур за смену, совместимость с другими процедурами. Множество параметров всех процедур образует матрицу Q размерности $s \times t$.

Оздоровительная программа представляет собой упорядоченное подмножество процедур с указанием их количества и периодичности. Обозначим множество всех допустимых оздоровительных программ как $\Pi = \{\pi_1, \pi_2, \dots, \pi_u\}$, где u – число программ. Каждая программа $\pi_k \in \Pi$ определяется как вектор назначений $\pi_k = (a_{k1}, a_{k2}, \dots, a_{ks})$, где $a_{kj} \in \{0, 1, 2, \dots, a_{\text{max}}\}$ – количество назначений процедуры c_j в рамках программы π_k (нулевое значение означает, что процедура не назначена, a_{max} – максимальное допустимое количество процедур данного типа за смену).

Множество программ формируется на основе медицинских стандартов и экспертных знаний врачей ДСОК и может корректироваться при изменении материально-технической базы учреждения.

Модель профиля здоровья ребенка

Профиль здоровья ребёнка p_i является центральной сущностью математической модели персонализации. Его построение осуществляется на этапе первичного обследования при поступлении в ДСОК и обновляется по итогам промежуточных и итогового обследований.

Для обеспечения сопоставимости разнородных признаков в рамках единой модели все компоненты вектора p_i приводятся к нормализованному виду. Номинальные признаки кодируются с помощью one-hot кодирования: каждому значению номинального признака $f_j \in F^{\text{nom}}$ ставится в соответствие бинарная переменная. Порядковые признаки преобразуются в числовые значения на основе экспертно заданной шкалы. Количественные непрерывные признаки подвергаются z-нормализации:

$$p'_{ij} = (p_{ij} - \mu_j) / \sigma_j, \quad (1)$$

где μ_j и σ_j – выборочное среднее и стандартное отклонение j -го признака, вычисленные по данным всех детей, проходивших оздоровление в данном ДСОК за предшествующий период (не менее трёх смен). Бинарные признаки остаются без преобразования.

Нормализованный профиль здоровья ребёнка обозначим как $\hat{p}_i$. Размерность вектора $\hat{p}_i$ после кодирования номинальных признаков может существенно превышать исходное число признаков m , однако это не создаёт вычислительных проблем при средних значениях n и m , характерных для ДСОК. Параметры нормализации (μ_j , σ_j) рассчитываются по историческим данным учреждения. Такой подход гарантирует, что преобразование остаётся стабильным при поступлении новых детей, а профили, полученные в разных сменах, оказываются корректно сопоставимыми между собой.

Вектор противопоказаний $r_i = (r_{i1}, r_{i2}, \dots, r_{is})$ дополняет профиль здоровья. Компонента r_{ij} принимает значение 1, когда процедура c_j противопоказана ребёнку d_i по состоянию здоровья, и 0 – во всех остальных случаях. Формирование этого вектора выполняется врачом на основании диагноза, анамнеза и результатов обследования. Для процесса подбора программы он играет роль жёсткого ограничения: ребёнку не может быть назначена ни одна программа, в которую входит хотя бы одна противопоказанная для него процедура.

Задача подбора оздоровительной программы

Подбор оздоровительной программы для ребёнка d_i сводится к выбору из множества Π такой программы π_k , которая наилучшим образом отвечает профилю здоровья $\hat{r}_i$ и при этом не нарушает медицинских ограничений. Формально требуется найти программу π^* , максимизирующую функцию соответствия (fitness function) между профилем ребёнка и параметрами программы, при условии отсутствия нарушений противопоказаний и правил совместимости процедур.

Введём функцию соответствия $\varphi(\hat{r}_i, \pi_k) \rightarrow \mathbb{R}$, которая количественно оценивает степень адекватности назначения программы π_k ребёнку с профилем $\hat{r}_i$. Чем выше значение функции соответствия, тем более обоснованным считается назначение данной программы. Тогда задача подбора формулируется как оптимизационная:

$$\pi^* = \arg \max_{\{\pi_k \in \Pi\}} \varphi(\hat{r}_i, \pi_k), \text{ при ограничениях:} \quad (2)$$

1. Отсутствие противопоказанных процедур: $\forall j : a_{kj} > 0 \Rightarrow r_{ij} = 0$. Это условие гарантирует, что ни одна назначенная процедура не входит в перечень противопоказаний ребёнка.

2. Совместимость процедур: $\forall j, l : a_{kj} > 0 \wedge a_{kl} > 0 \Rightarrow \text{compat}(c_j, c_l) = 1$, где $\text{compat}(c_j, c_l)$ – функция совместимости процедур, принимающая значение 1, если процедуры c_j и c_l совместимы, и 0 в противном случае.

Правила совместимости формализуются на основе медицинских регламентов и экспертных знаний.

3. Возрастные ограничения: $\forall j : a_{kj} > 0 \Rightarrow \text{age_min}(c_j) \leq \text{age}(d_i) \leq \text{age_max}(c_j)$, где age_min и age_max – минимальный и максимальный возраст назначения процедуры c_j .

4. Ограничение по группе здоровья: $\forall j : a_{kj} > 0 \Rightarrow \text{group}(c_j) \in \text{health_group}(d_i)$, где $\text{group}(c_j)$ – множество групп здоровья, для которых разрешена процедура c_j .

Множество программ, удовлетворяющих всем четырём ограничениям для данного ребёнка d_i , образует допустимое множество $\Pi_i \subseteq \Pi$. Если допустимое множество пусто, что на практике означает, что ни одна из стандартных программ не подходит ребёнку (например, при редком сочетании противопоказаний), врач формирует индивидуальную программу вручную, и данная программа добавляется в Π для последующего использования.

Оптимизационная постановка и вычислительная сложность

Задача подбора оздоровительной программы относится к классу задач дискретной оптимизации с ограничениями. Число допустимых программ u в типичном ДСОК составляет от 15 до 40 (по числу профилей оздоровления, утверждённых приказом главного врача), а число признаков m после кодирования – от 30 до 80. При указанных размерностях полный перебор допустимого множества Π_i является вычислительно тривиальной задачей: для каждого ребёнка требуется вычислить $\varphi(\hat{r}_i, \pi_k)$ для $|\Pi_i|$ программ и выбрать максимум, что требует $O(|\Pi_i| \cdot m)$ операций.

Однако при расширении постановки – например, при формировании программы не из фиксированного набора шаблонов, а путём комбинаторного выбора процедур из множества S – задача становится существенно более сложной. В этом случае число возможных комбинаций процедур экспоненциально зависит от s (числа доступных процедур), и

полный перебор становится непрактичным уже при $s > 25$. В данной работе принят шаблонный подход, при котором система ранжирует заранее сформированные программы из множества Π , что обеспечивает вычислительную эффективность и одновременно медицинскую корректность (каждая программа в Π прошла экспертизу). Расширение до комбинаторного формирования программ рассматривается как перспективное направление развития.

Функция соответствия профиля ребенка и параметров программы

Центральным элементом математической модели является функция соответствия $\varphi(\hat{p}_i, \pi_k)$, количественно оценивающая степень адекватности назначения программы π_k ребёнку с профилем $\hat{p}_i$. Функция соответствия строится как аддитивная свёртка нескольких компонент, каждая из которых отражает отдельный аспект соответствия.

Первая компонента – целевое соответствие $\varphi_{\text{target}}(\hat{p}_i, \pi_k)$ – оценивает, насколько программа направлена на решение проблем, выявленных у ребёнка. Вектор целевых показателей $\tau_k = (\tau_{k1}, \tau_{k2}, \dots, \tau_{km})$ описывает каждую программу π_k . Компонента $\tau_{kj} \in \{0, 1\}$ показывает, направлена ли программа на коррекцию j -го показателя здоровья: единица означает, что направлена, ноль – что нет.

Для ребёнка строится вектор потребностей $v_i = (v_{i1}, v_{i2}, \dots, v_{im})$. Здесь $v_{ij} = 1$, если j -й показатель отклонился от нормы и нуждается в коррекции, и $v_{ij} = 0$ во всех остальных случаях. Отклонение от нормы устанавливается по возрастным нормативам: для количественных показателей оно определяется выходом за границы стандартного коридора ($\mu \pm 1.5\sigma$), для категориальных – принадлежностью к III–V группам здоровья либо наличием конкретного диагноза.

Расчёт целевого соответствия выполняется через взвешенное скалярное произведение:

$$\varphi_{target}(\hat{p}_i, \pi_k) = \frac{\sum_{j=1}^m w_j \cdot v_{ij} \cdot \tau_{kj}}{\sum_{j=1}^m w_j \cdot v_{ij}} \quad (3)$$

Вес j -го показателя w_j задаётся экспертом (врачом) исходя из его клинической значимости, после чего веса нормируются так, чтобы выполнялось условие $\sum_j w_j = 1$. Наличие знаменателя позволяет привести значение целевого соответствия к диапазону $[0, 1]$: единица достигается в ситуации, когда программа перекрывает все выявленные у ребёнка потребности, ноль – когда не перекрывает ни одной.

Профильное сходство $\varphi_{profile}(\hat{p}_i, \pi_k)$ – вторая компонента – отвечает на вопрос, насколько ребёнок близок к типичному пациенту, для которого эта программа предназначалась. С каждой программой π_k связан эталонный профиль пациента $\tilde{\pi}_k = (\tilde{\pi}_{k1}, \dots, \tilde{\pi}_{km})$, который получается как центроид (средний вектор) профилей тех детей, у которых данная программа ранее показала высокую результативность. В качестве меры профильного сходства используется косинусная мера:

$$\varphi_{profile}(\hat{p}_i, \pi_k) = \frac{(\hat{p}_i \cdot \tilde{\pi}_k)}{(\|\hat{p}_i\| \cdot \|\tilde{\pi}_k\|)} \quad (4)$$

Выбор косинусной меры вместо евклидова расстояния не случаен. Косинусная мера нечувствительна к масштабу векторов и реагирует на направление отклонений, а не на их абсолютную величину. Для профилей здоровья это существенно: значение имеют не столько сами числовые значения показателей, сколько их соотношение между собой и характер структуры отклонений от нормы.

Историческая результативность $\varphi_{history}(\pi_k)$ – третья компонента – отражает, насколько эффективной программа π_k оказалась для детей, проходивших её ранее. Обозначим через N_k число пациентов, завершивших программу π_k в предшествующих сменах, а через e_k – долю тех из них, у кого по итогам программы зафиксирована положительная

динамика по ключевым показателям здоровья (улучшение хотя бы по одному значимому показателю). Тогда:

$$\varphi_history(\pi_k) = e_k = N_k^+ / N_k \quad (5)$$

где N_k^+ – число детей с положительной динамикой. Для вновь созданных программ, по которым ещё нет данных о результативности ($N_k = 0$), устанавливается $\varphi_history(\pi_k) = 0.5$ (нейтральное значение, отражающее отсутствие статистики). По мере накопления данных историческая результативность уточняется, обеспечивая реализацию обратной связи: программы, демонстрирующие высокую эффективность, получают преимущество при рекомендациях.

Интегральная функция соответствия и веса компонент

Итоговая функция соответствия формируется как взвешенная сумма трёх компонент:

$$\varphi(\hat{p}_i, \pi_k) = \alpha \cdot \varphi_target(\hat{p}_i, \pi_k) + \beta \cdot \varphi_profile(\hat{p}_i, \pi_k) + \gamma \cdot \varphi_history(\pi_k) \quad (6)$$

где α, β, γ – весовые коэффициенты, удовлетворяющие условиям $\alpha + \beta + \gamma = 1$ и $\alpha, \beta, \gamma \geq 0$. Выбор весов определяет относительную значимость каждого аспекта соответствия и может варьироваться в зависимости от приоритетов конкретного ДСОК.

По умолчанию предлагаются следующие значения: $\alpha = 0.45$, $\beta = 0.30$, $\gamma = 0.25$. Обоснование данного распределения состоит в том, что целевое соответствие (покрытие выявленных проблем) является наиболее клинически значимым критерием и должно иметь наибольший вес. Профильное сходство обеспечивает учёт индивидуальных особенностей ребёнка, выходящих за рамки формальных показаний, но его вклад вторичен по отношению к целевому назначению. Историческая результативность, хотя и представляет ценность для адаптации рекомендаций, на начальных этапах эксплуатации системы (при малом N_k) не может быть основным критерием и потому получает наименьший вес.

Таблица 1. Весовые коэффициенты функции соответствия и их обоснование

Компонента	Обозначение	Вес по умолчанию	Обоснование
Целевое соответствие	α	0.45	Покрытие выявленных проблем здоровья – основной критерий назначения
Профильное сходство	β	0.30	Учёт индивидуальных особенностей, не формализованных в показаниях
Историческая результативность	γ	0.25	Адаптация на основе накопленного опыта, вторична при малой выборке

Весовые коэффициенты будут доступны для настройки администратором планируемой к разработке информационной системы через интерфейс конфигурации, что позволяет учреждению адаптировать модель под свою специфику. Например, ДСОК с многолетней историей и большим объёмом накопленных данных может увеличить вес γ для более активного использования обратной связи.

Учет медицинских ограничений и противопоказаний

Механизм учёта противопоказаний реализован как двухуровневая система фильтрации. На первом уровне – жёсткая фильтрация – из множества допустимых программ Π_i исключаются все программы, содержащие хотя бы одну противопоказанную процедуру (ограничение 1 из раздела 2.1.3). Данная фильтрация является безусловной и не может быть overridden алгоритмом: если процедура противопоказана, она не назначается ни при каком значении функции соответствия. Реализация этого правила гарантирует безопасность рекомендаций и соответствие принципу «не навреди».

На втором уровне – мягкая фильтрация – учитываются правила совместимости процедур (ограничение 2), возрастные ограничения (ограничение 3) и ограничения по группе здоровья (ограничение 4). Программы, нарушающие хотя бы одно из этих ограничений, исключаются из Π_i , однако в отличие от жёсткой фильтрации, врач имеет возможность просмотреть полный ранжированный список программ (с пометкой о нарушенных ограничениях) и при наличии веских клинических оснований назначить программу вручную, подтвердив своё решение электронной подписью.

Матрица совместимости процедур C_compat размерности $s \times s$ задаётся экспертно и хранится в базе данных системы. Элемент $C_compat[j][l] = 1$, если процедуры c_j и c_l совместимы, и 0 в противном случае. Совместимость определяется на основе физиотерапевтических правил: например, некоторые виды электролечения несовместимы с теплотечением в один день, а грязевые процедуры не назначаются одновременно с соляной пещерой. Формально ограничение совместимости для программы π_k проверяется следующим образом: для всех пар $(j, l) : a_{kj} > 0 \wedge a_{kl} > 0$ должно выполняться $C_compat[j][l] = 1$. Вычислительная сложность проверки совместимости для одной программы составляет $O(s^2)$, что при типичных значениях s (30–80) не создаёт проблем производительности.

Ранжирование программ и формирование рекомендаций

Алгоритм формирования рекомендаций для каждого ребёнка d_i работает следующим образом. Сначала из общего множества программ Π выделяется допустимое подмножество Π_i – для этого последовательно применяются ограничения 1–4. Затем в пределах этого подмножества для каждой программы $\pi_k \in \Pi_i$ рассчитывается значение функции соответствия $\varphi(\hat{r}_i, \pi_k)$. После этого программы выстраиваются в порядке убывания φ , и врачу предъявляется ранжированный список: рядом с каждой программой

указывается значение ϕ и приводится расшифровка вклада трёх компонент – целевого соответствия, профильного сходства и исторической результативности.

К рекомендациям предъявляется жёсткое требование интерпретируемости: врач должен видеть, чем обусловлен выбор системы. С этой целью к каждой рекомендации формируется текстовое пояснение. Например: «Программа "Дыхание" рекомендована: целевое соответствие 0.87 (покрытие 6 из 7 выявленных проблем), профильное сходство 0.72 (близость к эталонному профилю пациентов данной программы), историческая результативность 0.81 (81 % предыдущих пациентов показали положительную динамику)». Такой формат позволяет врачу принять информированное решение: согласиться с рекомендацией, выбрать другую программу из ранжированного списка или назначить программу вручную.

Финальное назначение всегда остаётся за врачом. Система выступает в роли инструмента поддержки принятия решений, а не автоматического назначателя. Данное решение продиктовано как медицинскими этическими нормами (ответственность за назначение несёт врач), так и практическими соображениями: алгоритмическая рекомендация может не учитывать нюансы, не формализованные в профиле здоровья (психоэмоциональное состояние, мотивацию ребёнка, пожелания родителей).

Описательная статистика и первичный анализ данных

Модуль статистического анализа информационной системы реализует комплекс методов, последовательно углубляющих понимание данных о здоровье детей от описания к выявлению закономерностей и оценке динамики. Базовым уровнем является описательная статистика, обеспечивающая характеристику совокупности детей, проходящих оздоровление в ДСОК.

Для количественных непрерывных признаков (рост, масса тела, индекс массы тела, результаты функциональных проб) вычисляются: выборочное среднее $\bar{x} = (1/n) \sum_i x_i$, выборочное стандартное отклонение $s = \sqrt{[(1/(n-1)) \sum_i (x_i - \bar{x})^2]}$, медиана Me , первый и третий квартили Q_1 и Q_3 , межквартильный размах $IQR = Q_3 - Q_1$. Для выявления аномальных наблюдений применяется правило $1.5 \times IQR$: значение признаётся выбросом, если оно выходит за границы $[Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR]$. Выбросы не исключаются из анализа автоматически, но помечаются для внимания врача, поскольку могут отражать как ошибки ввода, так и клинически значимые крайние значения.

Для категориальных признаков (группа здоровья, диагноз, пол) вычисляются абсолютные и относительные частоты. Результаты представляются в виде таблиц частотного распределения и визуализируются в форме столбчатых диаграмм и круговых диаграмм. Система автоматически формирует сводную таблицу описательной статистики по каждой смене, по произвольному периоду и в разрезе групп здоровья, что позволяет врачу и администратору ДСОК оперативно получать характеристику контингента.

Важным элементом описательной статистики является оценка нормальности распределения количественных признаков, поскольку от неё зависит выбор параметрических или непараметрических методов в последующем анализе. Для проверки нормальности используется критерий Шапиро-Уилка, рекомендуемый для выборок объёмом от 20 до 5000 наблюдений, что соответствует типичным размерам выборок в ДСОК. При отклонении гипотезы нормальности ($p < 0.05$) система автоматически переключается на непараметрические методы анализа.

Корреляционный анализ факторов выздоровления

Корреляционный анализ применяется для выявления статистически значимых связей между показателями здоровья детей и характеристиками

оздоровительных программ. Его результаты служат двум целям: во-первых, они обосновывают выбор факторов, включаемых в модель подбора программ, во-вторых, позволяют оценить структуру взаимосвязей в самих данных.

Если пара количественных признаков (x, y) имеет распределение, близкое к нормальному, для неё рассчитывается коэффициент корреляции Пирсона:

$$r_{xy} = \Sigma_i (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{[\Sigma_i (x_i - \bar{x})^2 \cdot \Sigma_i (y_i - \bar{y})^2]} \quad (7)$$

Когда же признаки являются порядковыми либо количественными, но с отклонением от нормальности, обращаются к коэффициенту ранговой корреляции Спирмена:

$$\rho_{xy} = 1 - (6 \Sigma_i d_i^2) / (n(n^2 - 1)) \quad (8)$$

Здесь d_i обозначает разность рангов i -го наблюдения по переменным x и y . Что касается номинальных признаков, то для оценки связи между ними используется коэффициент V Крамера. Он строится на статистике хи-квадрат, нормируется в диапазон $[0, 1]$ и интерпретируется как мера силы связи.

Статистическая значимость корреляции оценивается путём проверки нулевой гипотезы $H_0: \rho = 0$. Для коэффициента Пирсона используется t -статистика $t = r\sqrt{(n - 2) / (1 - r^2)}$ с $(n - 2)$ степенями свободы. Для коэффициента Спирмена применяются табличные критические значения или аппроксимация t -распределением. Уровень значимости по умолчанию установлен $\alpha = 0.05$, однако система поддерживает настройку порога (0.01, 0.05, 0.10) в зависимости от требуемой строгости.

Результаты корреляционного анализа представляются в виде корреляционной матрицы с цветовой кодировкой силы и направления связей, а также в виде таблицы статистически значимых корреляций с указанием коэффициента, p -значения и объёма выборки. Интерпретация результатов сопровождается предупреждением о том, что корреляционная

связь не означает причинно-следственную зависимость, и для содержательных выводов необходимо привлечение экспертных знаний врача.

Регрессионный анализ результативности оздоровительных программ

Регрессионный анализ используется для построения модели, описывающей зависимость результативности оздоровления от характеристик ребёнка и параметров назначенной программы. В качестве зависимой переменной (отклика) выступает величина изменения ключевого показателя здоровья, например $\Delta\text{ИМТ} = \text{ИМТ}_{\text{выписка}} - \text{ИМТ}_{\text{поступление}}$ или $\Delta\text{Руфье} = \text{Руфье}_{\text{выписка}} - \text{Руфье}_{\text{поступление}}$. Независимыми переменными (предикторами) являются компоненты профиля здоровья ребёнка и признаки назначенной программы.

Модель множественной линейной регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (9)$$

где Y – вектор значений отклика, $X_1, \dots, X_p$ – предикторы, $\beta_0, \dots, \beta_p$ – оцениваемые коэффициенты, ε – вектор случайных ошибок. Оценка коэффициентов осуществляется методом наименьших квадратов (МНК), минимизирующим сумму квадратов остатков: $\hat{\beta} = \arg \min \sum_i (y_i - \hat{y}_i)^2$.

Качество регрессионной модели оценивается по нескольким критериям. Коэффициент детерминации $R^2 = 1 - \text{SS}_{\text{res}} / \text{SS}_{\text{tot}}$ показывает долю дисперсии отклика, объяснённую моделью. Скорректированный $R^2_{\text{adj}} = 1 - (1 - R^2)(n - 1)/(n - p - 1)$ учитывает число предикторов и штрафует за их избыточность. Статистическая значимость модели в целом проверяется с помощью F-критерия: $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$. Статистическая значимость отдельных коэффициентов проверяется t-критерием: $H_0: \beta_j = 0$.

Обязательным этапом является проверка предпосылок МНК: нормальность остатков (критерий Шапиро-Уилка на векторе остатков),

гомоскедастичность (тест Бройша-Пагана на наличие гетероскедастичности), отсутствие мультиколлинеарности (VIF-фактор инфляции дисперсии, порог $VIF < 10$). При нарушении предпосылок применяются робастные методы оценивания (взвешенный МНК) или модель переформулируется (логарифмирование отклика, исключение мультиколлинеарных предикторов).

Интерпретация регрессионной модели в контексте ДСОК имеет практическое значение: значимые коэффициенты β_j показывают, какие факторы (возраст, группа здоровья, набор процедур) статистически значимо влияют на результативность оздоровления и в каком направлении.

Оценка динамики показателей здоровья

Оценка динамики показателей здоровья является ключевым элементом аналитического модуля, позволяющим определить, произошло ли статистически значимое улучшение или ухудшение показателей по итогам оздоровления. Для каждого ребёнка d_i и каждого количественного показателя $f_j \in F^{\text{cont}}$ фиксируются два значения: «при поступлении» x_{ij_in} и «при выписке» x_{ij_out} . Разность $\Delta x_{ij} = x_{ij_out} - x_{ij_in}$ образует выборку связанных (парных) наблюдений.

Для проверки гипотезы о наличии систематического сдвига используется парный t-критерий Стьюдента при условии нормальности распределения разностей (проверяемой критерием Шапиро-Уилка) или непараметрический критерий Уилкоксона для связанных выборок при нарушении нормальности. Нулевая гипотеза $H_0: \mu_{\Delta} = 0$ (среднее изменение равно нулю) проверяется против двусторонней альтернативы $H_1: \mu_{\Delta} \neq 0$. Уровень значимости $\alpha = 0.05$.

Кроме статистической значимости, рассчитывается величина эффекта (effect size). Она нужна для того, чтобы оценить, имеют ли выявленные изменения практический смысл. Применительно к парному t-критерию в роли такой меры выступает коэффициент Коэна:

$$d = \bar{x}_{\Delta} / s_{\Delta} \quad (10)$$

Здесь $\bar{x}_{\Delta}$ – среднее значение разностей, s_{Δ} – стандартное отклонение этих разностей. Само значение d соотносят с качественной шкалой: если $d < 0.2$, эффект считается пренебрежимо малым; при $0.2 \leq d < 0.5$ – малым; при $0.5 \leq d < 0.8$ – средним; наконец, когда $d \geq 0.8$, говорят о большом эффекте. Подобное разграничение даёт врачу возможность отделить статистически значимое, но клинически незначительное улучшение (скажем, $d = 0.15$ на большой выборке) от улучшения, значимого по существу ($d = 0.9$).

Итоги оценки динамики сводятся в единую таблицу. По каждому показателю в ней приводятся: средние значения «до» и «после», среднее изменение $\bar{x}_{\Delta}$ и стандартное отклонение этого изменения s_{Δ} , значение t -статистики либо Z -статистики Уилкоксона, p -значение, коэффициент Коэна d и качественная характеристика величины эффекта.

Метрики качества рекомендательных алгоритмов

Для оценки качества рекомендаций персонализированного подбора программ используется комплекс метрик, традиционно применяемых в рекомендательных системах и адаптированных к предметной области детского оздоровления.

Точность рекомендаций (Recommendation Accuracy) оценивается путём сравнения рекомендаций системы с фактическими назначениями врачей, обладающих высокой квалификацией и опытом работы в ДСОК. Для каждого ребёнка d_i формируется множество фактически назначенных программ A_i (как правило, $|A_i| = 1$) и множество программ, рекомендованных системой на позициях с 1 по K – $R_i(K) = \{\pi_1, \pi_2, \dots, \pi_K^*\}$. Метрика Hit Rate@K определяется как доля случаев, в которых фактически назначенная программа оказалась среди K рекомендаций:

$$HR@K = (1/n) \sum_{i=1}^n \mathbb{1}[A_i \cap R_i(K) \neq \emptyset] \quad (11)$$

где $\mathbb{1}[\cdot]$ – индикаторная функция. Значение $HR@K = 1$ означает, что во всех случаях назначение врача совпало хотя бы с одной из K рекомендаций системы. В данной работе используются значения $K = 1, 3, 5$, что позволяет оценить качество как наиболее релевантной рекомендации ($K = 1$), так и качества ранжирования при расширении списка.

Средняя обратная позиция (Mean Reciprocal Rank, MRR) учитывает не только факт совпадения, но и позицию фактически назначенной программы в ранжированном списке:

$$MRR = (1/n) \sum_{i=1}^n 1 / rank_i \quad (12)$$

где $rank_i$ – позиция фактически назначенной программы в ранжированном списке рекомендаций для i -го ребёнка. $MRR = 1$ означает, что все назначения врачей совпали с первыми рекомендациями системы.

Нормализованная дисконтированная кумулятивная прибыль ($nDCG@K$) оценивает качество ранжирования с учётом убывания релевантности по позициям: чем выше релевантная программа в списке, тем выше вклад в метрику. Формально:

$$nDCG@K = \frac{(1/n) \sum_i DCG_i@K}{IDCG_i@K} \quad (13)$$

где $DCG_i@K = \sum_{j=1}^K rel_{ij} / \log_2(j + 1)$, rel_{ij} – релевантность j -й рекомендации для i -го ребёнка (1 – совпадение с назначением врача, 0 – несовпадение), $IDCG_i@K$ – идеальное значение DCG при наилучшем ранжировании.

Статистическая значимость и методика проведения эксперимента

Для обеспечения корректности выводов о качестве рекомендаций необходимо оценить статистическую значимость наблюдаемых различий между рекомендациями системы и назначениями врачей. Поскольку объём выборки в ДСОК ограничен (100–500 наблюдений за смену), применяются методы, устойчивые к малым выборкам.

Для сравнения HR@K системы с базовым уровнем (случайный выбор программы из допустимого множества) используется биномиальный критерий: если HR@K при случайном выборе равна p_0 , а система достигла HR@K = p_1 на выборке объёма n , то проверяется гипотеза H_0 : $p_1 = p_0$ против H_1 : $p_1 > p_0$. Биномиальный критерий выбирается потому, что HR@K по своей природе является частотой бинарных исходов (попадание / промах).

Для сравнения качества ранжирования (MRR, nDCG) между различными конфигурациями модели (например, при разных весовых коэффициентах α , β , γ) используется критерий Уилкоксона для связанных выборок, поскольку значения метрик вычисляются на одних и тех же детях и образуют парные наблюдения.

Методика проведения эксперимента включает следующие этапы. Формируется тестовая выборка из данных о детях, прошедших оздоровление в предшествующих сменах (не менее 100 записей), для которых известны профили здоровья, фактически назначенные программы и результаты оздоровления. Для каждого ребёнка из тестовой выборки система формирует рекомендации, которые сравниваются с фактическими назначениями. Вычисляются метрики HR@1, HR@3, HR@5, MRR, nDCG@5. Проводится сравнение с базовыми моделями: случайный выбор из допустимого множества и выбор наиболее популярной программы (по частоте назначения в обучающей выборке).

Метрики производительности модулей системы

Помимо качества рекомендаций, оценивается производительность программной реализации модулей системы. Для модуля персонализации измеряется время формирования рекомендаций для одного ребёнка и для группы детей (пакетный режим). Целевой показатель – время рекомендаций для одного ребёнка не более 200 мс, для группы из 500 детей – не более 5 с. Данные показатели обеспечивают комфортную работу

врача в интерактивном режиме и возможность оперативной обработки всех детей смены.

Для модуля статистического анализа измеряется время формирования описательной статистики, корреляционной матрицы, регрессионной модели и оценки динамики показателей. Целевой показатель – не более 3 с для каждого типа анализа на выборке до 1000 записей. Данные требования обусловлены характером использования: врач или администратор ожидает результат аналитического запроса в реальном времени.

Для модуля отчётности измеряется время генерации документов основных типов (санаторно-курортная карта, отчёт для ФСС, аналитическая справка). Целевой показатель – не более 5 с для одного документа и не более 30 с для комплекта отчётов по итогам смены. Сравнение проводится с временем ручного формирования аналогичных документов, которое по данным хронометража составляет от 15 до 45 минут на один документ и от 3 до 5 рабочих дней на комплект отчётов по смене.

Выводы

В ходе разработки математического аппарата формализована задача персонализированного подбора оздоровительных программ для детей в условиях ДСОК. Введена система множеств и обозначений, описывающая ключевые сущности предметной области: профили здоровья детей, параметры оздоровительных процедур, множество допустимых программ и вектор противопоказаний.

Разработана функция соответствия $\varphi(\hat{r}_i, \pi_k)$, являющаяся взвешенной аддитивной свёрткой трёх компонент: целевого соответствия (покрытие выявленных проблем здоровья), профильного сходства (косинусная мера близости профиля ребёнка к эталонному профилю программы) и исторической результативности (доля положительных исходов для

программы в предшествующих сменах). Обоснованы весовые коэффициенты компонент и возможность их настройки под специфику учреждения.

Описана двухуровневая система фильтрации, обеспечивающая безопасность рекомендаций: жёсткая фильтрация по противопоказаниям (безусловное исключение программ с противопоказанными процедурами) и мягкая фильтрация по совместимости, возрастным ограничениям и группам здоровья с возможностью экспертной коррекции.

Определён комплекс статистических методов, реализуемых в модуле аналитики: описательная статистика с проверкой нормальности, корреляционный анализ (Пирсон, Спирмен, V Крамера), множественная линейная регрессия с проверкой предпосылок МНК и робастным оцениванием, оценка динамики показателей здоровья с помощью парных критериев и вычислением величины эффекта по Коэну.

Определены метрики оценки качества рекомендаций (HR@K, MRR, nDCG@K), методика проведения эксперимента и целевые показатели производительности модулей системы. Установлены статистические процедуры проверки значимости наблюдаемых различий.

Разработанный математический аппарат составляет основу для проектирования и программной реализации планируемой к разработке в дальнейшем информационной системы.

ЛИТЕРАТУРА

1. Нелинейная математическая модель ценообразования продукции перерабатывающего предприятия / В. В. Степанов, Г. А. Аршинов, С. В. Лаптев, И. А. Мануйлов // Автоматизированные информационные и электроэнергетические системы : Материалы II Межвузовской научно-практической конференции, Краснодар, 07–09 сентября 2012 года / ФГБОУ ВПО КубГТУ. – Краснодар: Общество с ограниченной ответственностью "Издательский Дом - Юг", 2012. – С. 38-40. – EDN TCIENJ.

2. Лаптев, С. В. Постановка курса "Web-технологии в идентификации систем" / С. В. Лаптев // Качество современных образовательных услуг - основа конкурентоспособности вуза: сборник статей по материалам межфакультетской

учебно-методической конференции / Ответственный за выпуск М. В. Шаталова : Кубанский государственный аграрный университет, 2016. – С. 298-300. – EDN WBILOX.

3. Лаптев, С. В. Разработка информационных систем на базе web-технологий : учебное пособие / С. В. Лаптев, В. Н. Лаптев, Г. А. Аршинов. – Краснодар : Кубанский государственный аграрный университет им. И.Т. Трубилина, 2021. – 175 с. – ISBN 978-5-907430-34-1. – EDN QUGCPI.

4. Аршинов, Г. А. Анализ оборота капитала и цены на готовую продукцию в интегрированных объединениях АПК / Г. А. Аршинов, С. В. Лаптев, В. Г. Аршинов // Новые технологии. – 2018. – № 4. – С. 96-101. – EDN YXRFNJ.

5. Разработка адаптивной матрицы типовых знаний для инвестиционного управления АПК1 / В. Н. Лаптев, Г. А. Аршинов, С. В. Лаптев [и др.] // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2020. – № 164. – С. 36-54. – DOI 10.21515/1990-4665-164-003. – EDN LGPWRC.

6. Лаптев, С. В. Разработка аппаратно-программной системы для повышения эффективности контроля и управления работой сельскохозяйственных машин при выполнении полевых операций / С. В. Лаптев, В. С. Лаптева, С. А. Дельнов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2025. – № 211. – С. 22-41. – DOI 10.21515/1990-4665-211-003. – EDN YNZFOY.

7. Дискретная математика : Учебное пособие / С. В. Лаптев, В. Н. Лаптев, Г. А. Аршинов, А. Э. Сергеев. – Краснодар : Кубанский государственный аграрный университет, 2024. – 200 с. – ISBN 978-5-907816-69-5. – EDN PTYWZR.

REFERENCES

1. Nelinejnaja matematicheskaja model' cenoobrazovaniya produkcii pererabatyvajushhego predpriyatija / V. V. Stepanov, G. A. Arshinov, S. V. Laptev, I. A. Manujlov // Avtomatizirovannye informacionnye i jelektrojenergeticheskie sistemy : Materialy II Mezhvuzovskoj nauchno-prakticheskoj konferencii, Krasnodar, 07–09 sentjabrja 2012 goda / FGBOU VPO KubGTU. – Krasnodar: Obshhestvo s ogranichennoj otvetstvennost'ju "Izdatel'skij Dom - Jug", 2012. – S. 38-40. – EDN TCIEHJ.

2. Laptev, S. V. Postanovka kursa "Web-tehnologii v identifikacii sistem" / S. V. Laptev // Kachestvo sovremennyh obrazovatel'nyh uslug - osnova konkurentosposobnosti vuza: sbornik statej po materialam mezhfakul'tetskoj uchebno-metodicheskoj konferencii / Otvetstvennyj za vypusk M. V. Shatalova : Kubanskij gosudarstvennyj agrarnyj universitet, 2016. – S. 298-300. – EDN WBILOX.

3. Laptev, S. V. Razrabotka informacionnyh sistem na baze web-tehnologij : uchebnoe posobie / S. V. Laptev, V. N. Laptev, G. A. Arshinov. – Krasnodar : Kubanskij gosudarstvennyj agrarnyj universitet im. I.T. Trubilina, 2021. – 175 s. – ISBN 978-5-907430-34-1. – EDN QUGCPI.

4. Arshinov, G. A. Analiz oborota kapitala i ceny na gotovuju produkciju v integrirovannyh ob#edinenijah APK / G. A. Arshinov, S. V. Laptev, V. G. Arshinov // Novye tehnologii. – 2018. – № 4. – S. 96-101. – EDN YXRFNJ.

5. Razrabotka adaptivnoj matricy tipovyh znaniy dlja investicionnogo upravlenija APK1 / V. N. Laptev, G. A. Arshinov, S. V. Laptev [i dr.] // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2020. – № 164. – S. 36-54. – DOI 10.21515/1990-4665-164-003. – EDN LGPWRC.

6. Laptev, S. V. Razrabotka apparatno-programmnoj sistemy dlja povyshenija jeffektivnosti kontrolja i upravlenija rabotoj sel'skohozjajstvennyh mashin pri vypolnenii

polevyh operacij / S. V. Laptev, V. S. Lapteva, S. A. Del'nov // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2025. – № 211. – S. 22-41. – DOI 10.21515/1990-4665-211-003. – EDN YNZFOY.

7. Diskretnaja matematika : Uchebnoe posobie / S. V. Laptev, V. N. Laptev, G. A. Arshinov, A. Je. Sergeev. – Krasnodar : Kubanskij gosudarstvennyj agrarnyj universitet, 2024. – 200 s. – ISBN 978-5-907816-69-5. – EDN PTYWZR.