

Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования «Национальный  
исследовательский университет «Высшая школа экономики»

*На правах рукописи*

**Прокофьева Елизавета Сергеевна**

**СОВЕРШЕНСТВОВАНИЕ ПРОЦЕССА ОБСЛУЖИВАНИЯ В ГОРОДСКИХ  
ПОЛИКЛИНИКАХ НА ОСНОВЕ АДАПТИВНЫХ МОДЕЛЕЙ  
ПРОГНОЗИРОВАНИЯ ПОТОКОВ ПАЦИЕНТОВ**

**РЕЗЮМЕ**

**диссертации на соискание ученой степени  
кандидата компьютерных наук**

Москва- 2021

Диссертационная работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики»

**Научный руководитель:** Мальцева Светлана Валентиновна, доктор технических наук, профессор, Национальный исследовательский университет «Высшая школа экономики»

## **Тема диссертации**

Стремительно развивающиеся технологии анализа данных имеют колоссальное значение в здравоохранении. Существующий уровень автоматизации медицинского обслуживания позволяет накапливать и обрабатывать большие массивы данных и использовать их для решения оптимизационных задач. Кроме того, не только объем, но и разнообразие информации, используемой в здравоохранении, открывает новые возможности для продвинутой аналитики, в частности, больших данных. Таким образом, открытые данные Интернет-ресурсов, чаты пациентов и их отзывы о полученных услугах, электронные рецепты, обзоры аптек, ленты новостей социальных платформ, онлайн-консультации и многие другие источники данных становятся ресурсами для получения новых ценных знаний в области здравоохранения, способствующими развитию модели персонализированных услуг.

В рамках национального проекта по цифровизации экономики разработана и утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р Программа «Цифровая экономика Российской Федерации», в которой среди реализации отдельных направлений по отраслям в числе приоритетных указано здравоохранение. Одними из основных сквозных цифровых технологий Программы являются большие данные, нейротехнологии и искусственный интеллект. Следовательно, на сегодняшний день задачи по анализу и моделированию больших массивов данных городского здравоохранения с помощью технологий искусственного интеллекта представляет особую значимость для развития отраслевых решений в рамках цифровизации экономики, где данные являются ключевым фактором производства.

Важной областью применения технологий обработки данных в здравоохранении является совершенствование работы медицинских учреждений: эффективный график работы медицинского персонала, прогнозирование потока пациентов, планирование и распределение ресурсов, сокращение очередей и другие задачи. Современные аналитические технологии позволяют разрабатывать инструменты для принятия решений, в основе которых лежат эмпирические данные. Например, агрегированные данные фактических перемещений пациентов между медицинскими учреждениями и специалистами в рамках этих учреждений позволяют планировать загруженность ресурсов, обеспечивать высокий уровень доступности услуг и оптимизировать работу организации, исходя из реального

спроса на эти услуги. На основе таких данных активно развиваются информационные системы управления потоками пациентов.

Аналитика больших данных позволяет не только обнаруживать основные траектории перемещений в рамках лечебно-профилактических учреждений (ЛПУ), но также выявлять наиболее узкие места системы обслуживания в них для удовлетворения спроса населения на услуги при заданных экономических параметрах. Важным аспектом является возможность формирования прогнозных значений, как отправной точки для формирования альтернативной организационной модели оказания услуг с оптимально задействованными ресурсными единицами. Одним из важных требований к используемым прогнозным моделям является их адаптивность, так как данное требование обеспечивает учет влияния внешних факторов на входные потоки медицинской организации.

Международные системы здравоохранения продемонстрировали беспрецедентно растущие потребности в адаптивных методах обслуживания потоков пациентов в связи с распространением вируса SARS-CoV-2, признанного пандемией ВОЗ (Всемирная организация здравоохранения). Разработка данных методов позволит значительно повысить эффективность процесса оказания медицинских услуг и снизить расходы на его организацию, обеспечив доступность услуг населению.

Таким образом, разработка и применение технологий анализа данных в задачах анализа и прогнозирования потоков пациентов в ЛПУ является основой решения задач совершенствования процесса оказания услуг и ресурсного планирования, что определяет актуальность темы данного исследования.

### **Цель и задачи исследования**

Цель диссертационного исследования заключается в развитии методологии определения рациональной ресурсной структуры многопрофильной сервисной организации, обеспечивающей принятие управленческих решений путем разработки прогнозных моделей для удовлетворения спроса на услуги при имеющихся экономических ограничениях.

Для достижения цели исследования были поставлены следующие задачи:

1. Анализ существующих подходов и методов к моделированию и совершенствованию рациональной организационной структуры сервисных организаций с учетом особенностей входных потоков субъектов обслуживания.

2. Разработка и реализация метода анализа и типизации входного потока заявок на основе алгоритмов нечёткой кластеризации.
3. Разработка метода прогнозирования входного потока заявок, позволяющего учесть особенности изменения внешних факторов.
4. Разработка методики интеграции имитационной модели обслуживания в ЛПУ с методами анализа, типизации и прогнозирования входного потока заявок.

### **Научная новизна**

В данной работе впервые установлены следующие научные результаты, выносимые на защиту:

1. Оригинальный метод анализа и типизации входного потока заявок на основе нечеткой кластеризации в рамках теории аддитивной регуляризации тематических моделей.
2. Адаптивный метод прогнозирования входных потоков, обеспечивающий учет влияния внешних факторов на входные потоки ЛПУ.
3. Модернизированная модель обслуживания в ЛПУ на основе интеграции имитационной модели с разработанными методами анализа, типизации и прогнозирования входного потока заявок.

### **Практическая значимость**

Использование результатов данной работы на практике позволит добавить новые функциональные возможности в комплексы программного инструментария, поддерживающего работу лечебных учреждений. Также внедрении результатов исследования в учебный процесс при преподавании курса «Introduction to Big Data» на факультете бизнеса и менеджмента Национального исследовательского университета «Высшая школа экономики», кафедре инноваций и бизнеса в сфере информационных технологий, направление Бизнес- информатика, 3-й курс, 1 модуль, 2018-2019, 2019-2020. Также результаты исследования легли в основу докладов на лекциях и семинарах Зимней школы по бизнес-информатике для поступающих в магистратуру Национального исследовательского университета «Высшая школа экономики», в период с 2017-2020.

### **Методологическая база исследования**

Методы анализа данных, теория систем и системный анализ, теория графов, теория множеств, математическое моделирование, теория вероятностей и математическая статистики, методы анализа, моделирования и

мониторинга клинических путей пациентов, современные технологии программирования и имитационного моделирования.

**Апробация результатов исследования.** Основные результаты исследования были представлены на следующих конференциях и семинарах:

1. XXII Апрельская международная конференция по проблемам развития экономики и общества, 13-30 апреля 2021, Москва. Тема доклада: «Прогнозирование типизированного входного потока пациентов в многопрофильных медицинских организациях».
2. The International Conference on Information Systems (ICIS). 5<sup>th</sup> Annual Workshop on Big Data Applications, 13-16 December 2020, Hyderabad, India. Тема доклада: «Forecasting Heterogeneous Patient Flow through Big Data Application in Medical Facilities for Rational Staffing».
3. 22nd IEEE International Conference on Business Informatics (IEEE CBI 2020) June 22–24, 2020, Antwerp, Belgium. Тема доклада: «Driven Approach To Patient Flow Management And Resource Utilization In Urban Medical Facilities».
4. XXI Апрельская международная конференция по проблемам развития экономики и общества, 6-10 апреля 2020, Москва. Тема доклада: «Применение методов жесткой и нечеткой кластеризации для сегментации входного потока пациентов в городских поликлиниках».
5. Зимняя школа по бизнес-информатике для поступающих в магистратуру Национального исследовательского университета «Высшая школа экономики», 15-19 февраля 2020, пансионат «Солнечная поляна», Волково, Московская область. Тема доклада: «Анализ и визуализация медицинских данных в Python для прогнозирования посещаемости пациентов».
6. The International Conference on Information Systems (ICIS). Annual Workshop on Big Data Applications, 15-18 December 2019, Munich, Germany. Тема доклада: «Data-Driven Approaches for Efficient Patient Flow Segmentation in Polyclinics».
7. Ежегодный Саммит молодых ученых и инженеров «Большие вызовы для общества, государства и науки», ноябрь, 2019, Сириус, Сочи. Тема доклада: «Анализ и моделирование клинических путей пациентов для повышения качества обслуживания в городских поликлиниках».
8. 21st IEEE 21<sup>st</sup> International Conference on Business Informatics (IEEE CBI 2020), July 15-17, 2019, Moscow, Russia. Тема доклада:

«Application of Modern Data Analysis Methods to Cluster the Clinical Pathways in Urban Medical Facilities».

9. XX Апрельская международная конференция по проблемам развития экономики и общества, 9-12 апреля 2019, Москва. Тема доклада: «Применение современных методов анализа данных при моделировании клинических путей пациентов в городских медицинских учреждениях».
10. Russian-French Workshop in Big Data and Applications, 2017, Moscow, Russia. Тема доклада: «Modeling Demand for Services in the Healthcare Facilities Using Big Data Techniques».

### **Публикации стандартного уровня:**

1. Прокофьева Е.С., Мальцева С.В., Цю-жен-цин Д.А., Автоматизация управления ресурсами организации на основе имитационного моделирования с использованием прогнозирования входных потоков//Автоматизация. Современные технологии. 2021. Т. 75 № 6. С.272-280.
2. Prokofyeva E., Maltseva S., Tsiu-Zhen-Tsin D. Forecasting Heterogeneous Patient Flow through Big Data Application in Medical Facilities for Rational Staffing// International Conference on Information Systems 2020, Special Interest Group on Big Data Proceedings.
3. Prokofyeva E. S., Maltseva S. V., Fomichev N. Y., Kudryashov A. G. Data-Driven Approach To Patient Flow Management And Resource Utilization In Urban Medical Facilities// 22nd IEEE International Conference on Business Informatics (IEEE CBI 2020). P. 71-77.
4. Прокофьева Е.С., Зайцев Р.Д. Анализ клинических путей пациентов в медицинских учреждениях на основе методов жесткой и нечеткой кластеризации // Бизнес-информатика. 2020. Т. 14. № 1. С. 19–31.
5. Prokofyeva E.S., Zaytsev R.D., Maltseva S.V. Application of Modern Data Analysis Methods to Cluster the Clinical Pathways in Urban Medical Facilities// 21st IEEE International Conference on Business Informatics (IEEE CBI 2019). P. 75-83.
6. Prokofyeva E., Maltseva S. Data-Driven Approaches for Efficient Patient Flow Segmentation in Polyclinics// International Conference on Information Systems 2019, Special Interest Group on Big Data Proceedings.
7. Prokofyeva E.S., Zaytsev R.D., Maltseva S.V. The Demand for the Healthcare Services: the Opportunities of Big Data in Predicting Patient Flow//International Conference Information Systems 2017 Special Interest

## Содержание работы

Во введении обоснована актуальность исследования, поставлены цель, задачи, практическая и теоретическая значимость исследования, сформулированы научная новизна, достоверность и обоснованность полученных результатов. Также описана апробация результатов исследования

**Первая глава** посвящена текущему состоянию области, особенности моделирования системы обслуживания многопрофильных сервисных организаций, существующие подходы и методологическая база моделирования клинических путей как эффективного инструмента управления здравоохранением, поставлена основная задача исследования.

Важной областью технологий обработки данных в здравоохранении является оптимизация работы медицинских учреждений: эффективный график работы медицинского персонала, прогнозирование потока пациентов, планирование и распределение ресурсов, сокращение очередей и т.д. Современные аналитические технологии позволяют разрабатывать инструменты для принятия решений, в основе которых лежат эмпирические данные. Например, агрегированные данные фактических перемещений пациентов между медицинскими учреждениями и специалистами в рамках этих учреждений позволяют планировать загруженность ресурсов, обеспечивать высокий уровень доступности услуг и оптимизировать работу организации, исходя из реального спроса на эти услуги.

Разработка и внедрение клинических путей, или траекторий движения пациентов, является важным инструментом в управлении здравоохранением. В общем виде под клиническим путем пациента понимается траектория его движения при получении медицинской услуги в соответствующих учреждениях.

Для решения задачи исследования рассмотрена многоканальная система массового обслуживания (СМО) с ограниченным ожиданием.

Исследуется два вида входных потоков: статистический, составленный на основе данных прошлых периодов, и прогнозный. Заявка на обслуживание обладает неким типом, а обслуживающее устройство- производительностью. Заявка считается обслуженной, если узел произведет работу, окажет услугу по удовлетворению этой заявки.

В качестве ключевых критериев эффективности данной СМО с ограниченным ожиданием выбраны следующие характеристики обслуживания:

1. Среднее число заявок в очереди  $L_q$ ;
2. Среднее время пребывания заявки в очереди  $T_q$ ;
3. Интенсивность нагрузки канала обслуживания  $\varphi$ ;
4. Затраты на функционирование системы  $f$ .

Пусть затраты в моделируемой системе определяются выражением [1, 2]:

$$f = c_1 n + c_2 L_q(n, \rho) + c_3 \bar{k}_{\text{св}}(n) \rightarrow \min \quad (1)$$

где  $c_1$  – затраты на работу одного дополнительного канала обслуживания (провайдера услуги), отнесённые к единице времени,

$c_2$  – стоимостной эквивалент ожидания одной заявки в единицу времени,

$c_3$  – стоимостные потери от простоя одного канала в единицу времени.

Требуется определить рациональное число каналов обслуживания  $n$  при известных интенсивностях входного потока  $\lambda$  и потока обслуживания  $\mu$ , известных константах  $c_1, c_2, c_3$ . Также необходимо сравнить показатели функционирования системы  $\rho_0, Q, A, L_q, T_q, L_s, T_s$  для различных вариантов ограничений системы: на длину очереди  $m$ , интенсивности нагрузки канала обслуживания  $\varphi$ .

Ввиду сложности объекта управления и необходимости использования реальных данных большого объема задача не имеет точного решения, поэтому одним из подходов для исследования процессов функционирования систем массового обслуживания являются методы имитационного моделирования.

В данной работе впервые предлагается модернизированный метод типизации входного потока имитационной модели на базе нечёткой кластеризации в рамках теории аддитивной регуляризации тематических моделей (ARTM). Разработана схема включения прогнозных входных потоков в имитационную модель.

Входные потоки имитационной модели представлены временными рядами ввиду характерной структуры исходных данных медицинских организаций.

**Вторая глава** начинается с описания формального представления процесса медицинского обслуживания на основе теоретико-множественной модели.

Структура лечебного учреждения на основе теоретико-множественного описания может быть представлена как некоторая система обслуживания пациентов, которая в статике может быть описана следующей системой:

$$S_L = \{V, B, H, R\} \quad (2)$$

где  $V = \{v_j\}, j = 1, 2, \dots, M$  – множество пунктов оказания услуг в лечебном учреждении,  $v_j$  – код  $j$ -го пункта оказания услуг,  $M$  – количество пунктов оказания услуг; количество пунктов оказания услуг соответствует

количеству функциональных кабинетов лечебного учреждения. Перечень пунктов оказания услуг определяется Положением 10б организации оказания первичной медико-санитарной помощи взрослому населению.

Для элементов этого множества важным параметром является количество одновременно обслуживаемых пациентов  $V'$  и количество провайдеров, одновременно предоставляющих услуги  $V''$ . Зададим в виде множества параметров [3]:

$V' = \{v'_j\}, j = 1, 2, \dots, M$ , где  $v'_j$  – количество пациентов, которые могут одновременно обслуживаться в  $j$ -ом пункте оказания услуг.

$V'' = \{v''_j\}, j = 1, 2, \dots, M$ , где  $v''_j$  – количество провайдеров, которые могут одновременно принимать пациентов в  $j$ -ом пункте оказания услуг.

$H = \{h_q\}, q = 1, 2, \dots, Q$  – множество провайдеров услуг,  $h_q$  – код  $q$ -го провайдера,  $Q$  – количество провайдеров. Здесь провайдеры услуг – это врачи-специалисты, средний медицинский персонал, оказывающие медицинские услуги.

Множество связей  $R$  между объектами системы, заданными множествами  $V, B, H$ , может быть задано двумя матрицами –  $R'$  и  $R''$ .

Поскольку отдельный провайдер может оказывать несколько услуг, связь между услугами и провайдерами описывается матрицей:

$R' = \{r'_{lq}\}, l = 1, 2, \dots, L; q = 1, 2, \dots, Q$ , где  $r'_{lq} = 1$ , если в  $l$ -ый тип услуги может оказываться  $q$ -ым специалистом, 0 – в противном случае.

В условиях стабильного распределения провайдеров услуг по пунктам оказания услуг необходимо учитывать также связи между провайдерами и пунктами оказания услуг:

$R'' = \{r''_{ql}\}, q = 1, 2, \dots, Q; l = 1, 2, \dots, L$ , где  $r''_{ql} = 1$ , если  $q$ -ый провайдер оказывает услуги в  $l$ -ом пункте, 0 – в противном случае.

Очевидно, что конкретная услуга может оказываться в разных пунктах, например, терапевты могут принимать в нескольких кабинетах.

Динамика указанной системы определяется, в первую очередь, особенностями потока пациентов, а также динамикой состава объектов самой системы, например увольнения или временного отсутствия врачей или медперсонала, поломкой оборудования и т.п.

$P = \{p_i\}, i = 1, 2, \dots, N$  – множество пациентов,  $p_i$  –  $i$ -й пациент,

---

<sup>1</sup> Приказом Министерства здравоохранения и социального развития РФ от 15 мая 2012 г. N 543н "Об утверждении Положения об организации оказания первичной медико-санитарной помощи взрослому населению" (с изменениями и дополнениями) Приложение.

$N$  – количество пациентов. В действующей системе можно рассматривать  $N \rightarrow \infty$ , однако, целесообразно рассматривать  $N$ , соответствующее некоторому среднему количеству пациентов за определенный период, например, с учетом возможностей расписания работы специалистов.

$A = \{a_k\}, k = 1, 2, \dots, K$  – множество типов пациентов (в соответствии с диагнозами и видами клинических путей),  $a_k$  – код  $k$ -го типа пациента,  $K$  – количество типов пациентов.

Тип пациента является параметром для  $p_i$ , определяющим его клинический путь, поэтому можно определить множество параметров пациентов по этому виду классификации.

$A' = \{a'_i\}, i = 1, 2, \dots, N$  – множество типов пациентов из множества  $P$ ,  $a'_i$  – тип  $i$ -го пациента,  $a'_i \in A$ .

*Рекомендованный клинический путь пациента* – это рекомендованная на основании диагноза пациента упорядоченная последовательность медицинских услуг, оказываемых пациенту. В некоторых случаях при реализации допускается изменение последовательности оказания услуг пациентам. *Рекомендованный клинический путь* в рамках данного исследования формируется на основании алгоритмов ведения пациентов в Москве.

Введем следующие обозначения:

$S = \{s_i\}, i = 1, \dots, N$  – множество рекомендованных клинических путей пациентов,  $s_i$  – код клинического пути  $i$ -го пациента.

Зададим состав услуг, входящих в клинический путь  $s_i$ :

$A^i = \{a_n^i\}, n = 1, \dots, N^i$ , где  $a_n^i \in B$ .

Для клинического пути  $s_i$  можно задать номер начальной услуги,  $n_0 \in [1, N]$ , а также матрицу всех допустимых переходов между услугами, которые могут привести к разным их последовательностям:

$U^i = \{u_{lr}^i\}, l = 1, \dots, L; r = 1, \dots, L$ , где  $u_{lr}^i = 1$ , если клинический путь  $s_i$  допускает переход от услуги  $r$ -го типа к услуге  $l$ -го типа, 0 – в противном случае.

Клинический путь  $i$ -го пациента определяется на основании типа пациента,  $a'_i$ , а также его диагноза,  $a_i''$  если пациент относится к типу...

Наряду с рекомендованным клиническим путем пациента можно определить *реализованный клинический путь*, который определяется на основании реального прохождения пациентом рекомендованных обследований и специалистов.

Важно, что *реализованный* клинический путь может отличаться от *рекомендованного* ввиду ограничений ресурсов учреждения (например, при

наличии очередей), специфического поведенческого паттерна пациента, решения провайдера и других факторов.

Важными характеристиками реализованного клинического пути являются временные и стоимостные параметры пути, включая время прохождения всего пути, время ожидания того или иного исследования или специалиста. В случае платных медицинских услуг также стоимость реализации клинического пути и отдельных его этапов.

Таким образом можно задать *реализованный клинический путь* аналогично рекомендованному:

$S_p = \{s_{pi}\}, i = 1, \dots, N$  – множество реализованных клинических путей пациентов,  $s_{pi}$  - код реализованного клинического пути  $i$ -го пациента.

Зададим состав услуг, входящих в клинический путь  $s_{pi}$  :

$A_p^i = \{a_{pn}^i\}, n = 1, \dots, N_p^i$ , где  $a_{pn}^i \in B$ .

Для клинического пути  $s_{pi}$  можно задать порядковые номера услуг, коды специалистов, которые их оказывали, время ожидания услуги, время оказания услуги, стоимость услуги в следующих массивах:

С точки зрения обслуживания пациента система  $S_L = \{V, B, H, R\}$  должна обеспечивать минимизацию затрат времени на ожидание обслуживания и минимизацию стоимости обслуживания.

При этом система также должна обеспечивать минимизацию затрат на собственное функционирование, что предполагает, в частности, что врачи загружены в соответствии с нормативными показателями нагрузки, а также возможность предоставления сервисов по всем запросам клинических путей пациентов, что требует широкого спектра специалистов различного профиля.

Динамика работы лечебного учреждения также определяется неравномерностью потоков пациентов в разное время дня, а также вследствие различных нештатных ситуаций, например эпидемий, когда спрос на медицинские услуги резко повышается.

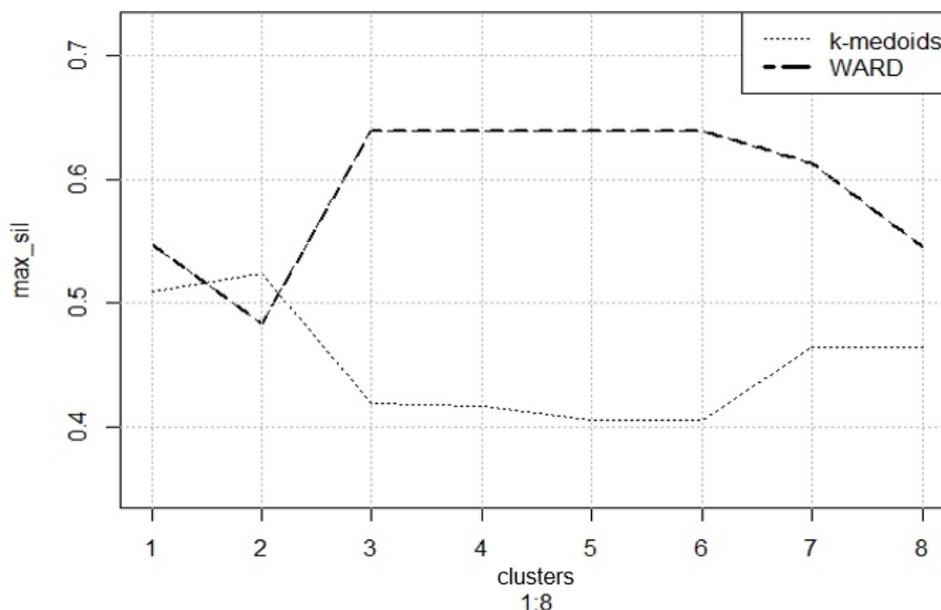
Типизация входного потока сводится к решению задачи кластеризации объектов. Для построения кластерной модели сравнивались два метода–  $k$ -medoids и иерархический агломеративный алгоритм с типом связи Уорда.

В качестве источника данных использован журнал событий голландской больницы, размещенный в открытом доступе. Журнал событий содержит 1143 маршрутов пациентов и 150291 событие [4].

После отбрасывания anomalously длинных путей верхний предел длины рассматриваемого пути полагался равным 26 событиям:  $Q50 + 3 \cdot (Q75 - Q50)$ , где  $Q50$  – медиана, а  $Q75$  соответствует 75% квантили. Для анализа данных был выбран язык программирования R. С помощью пакета Stringdist матрица

расстояний была построена методом Оса (мера расстояния Дамерау–Левенштейна).

Для поиска модели, в которой находится кластер с максимальным коэффициентом силуэта, сравнивались метод Уорда и *k-medoids* (рисунок 1), где по оси X расположены кластеры, по оси Y – значения силуэта.



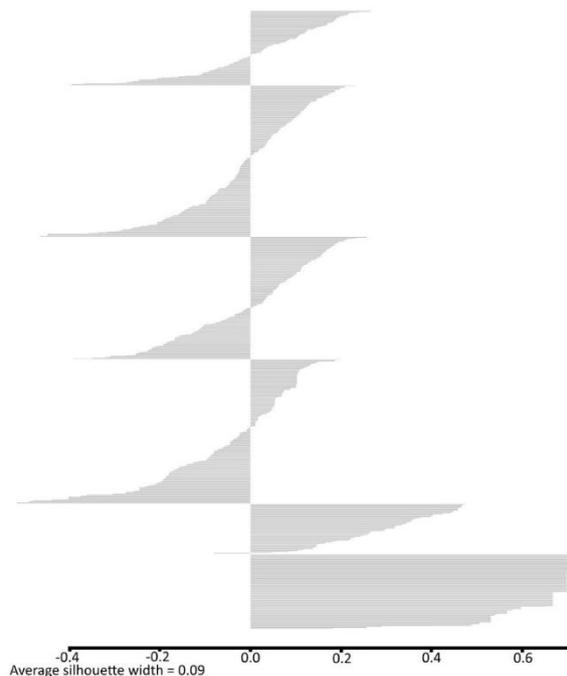
**Рисунок 1. Зависимость максимального коэффициента силуэта в модели от количества кластеров для методов *k-medoids* и Уорда**

В соответствии с проведенным анализом значений коэффициента был выбран метод Уорда и были выявлены тенденции полученных групп (рисунок 2). Кластеры с низким значением были исключены (рисунок 3), таким образом, в результате экспериментов были выбраны кластеры 5 и 6 с наибольшим коэффициентом силуэта (таблица 1).

Таблица 1.

**Значения коэффициента силуэта для кластеров**

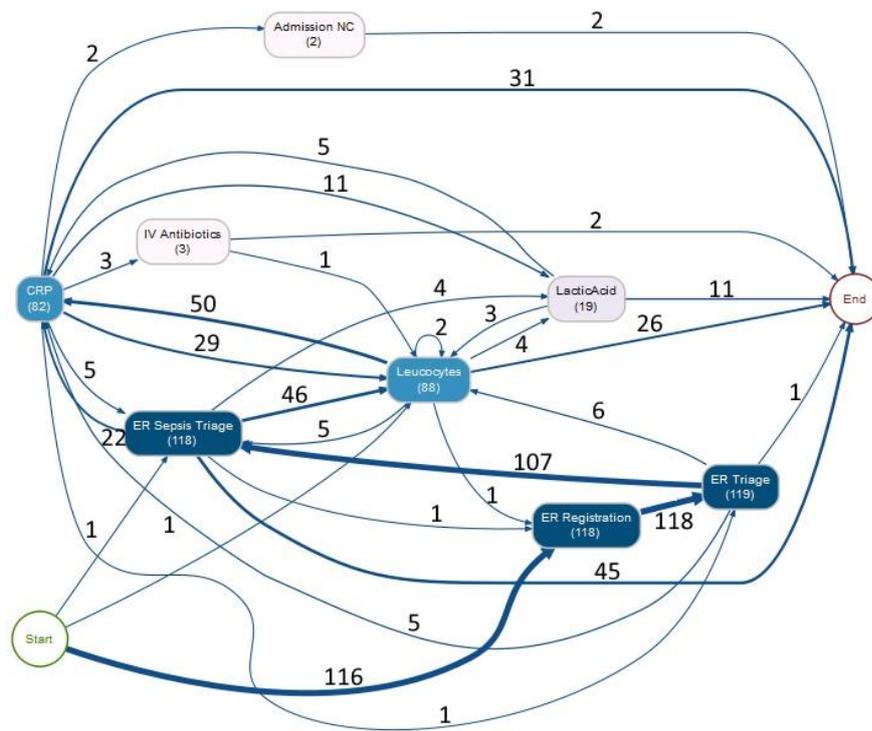
Номер кластера	Количество объектов	Значение коэффициента силуэта
1	118	0,02
2	239	-0,02
3	193	0,002
4	228	-0,05
5	79	0,29
6	118	0,64



**Рисунок 2. Коэффициенты силуэта шести результирующих кластеров**

После определения оптимального количества кластеров были отдельно сформированы процессные карты для полученных групп. Например, на процессной карте для кластера 5 в качестве узлов графа обозначены основные этапы клинического пути пациентов с диагнозом сепсис: старт, регистрация в соответствующем подразделении, прием антибиотиков и т.д. Дуги графа отображают переходы пациентов по этим этапам лечения, числа на дугах соответствуют количеству человек, совершающих данный переход между узлами. Более значимые траектории пациентов отмечены более широкими дугами. Таким образом, процессная карта позволяет быстро оценить наиболее загруженные маршруты медицинского учреждения. Кроме того, подобные карты могут быть интерпретированы медицинскими специалистами в разрезе сравнения их с принятыми медицинскими стандартами для выявления

перегруженных ресурсных единиц и дальнейшей реорганизации процесса обслуживания.



**Рисунок 3. Процессная карта для кластера 5**

Следующим этапом была нечеткая, или «мягкая», кластеризация исходных данных методами тематического моделирования, при которой путь пациента может относиться к нескольким шаблонам (кластерам-темам) с различными вероятностями.

Предложен модернизированный метод типизации входного потока объекта на основе нечёткой кластеризации в рамках теории аддитивной регуляризации тематических моделей (ARTM).

Построение тематической модели предполагает [5], что определены  $D$  – множество (коллекция) текстовых документов,  $W$  – словарь всех употребляемых в них терминов, каждый документ  $d \in D$  представляет собой последовательность  $n_d$  терминов  $(w_1, \dots, w_{n_d})$  словаря  $W$ .

Предполагается, что существует конечное множество тем  $T$ , и каждое употребление термина  $w$  в каждом документе  $d$  связано с некоторой темой  $t \in T$ , которая не известна. Коллекция документов рассматривается как множество троек  $(d, w, t)$  выбранных случайно и независимо из дискретного распределения  $p(d, w, t)$ , заданного на конечном множестве  $(D \times W \times T)$ .

Документы  $d \in D$  и термины  $w \in W$  являются наблюдаемыми переменными, тема  $t \in T$  является латентной (скрытой) переменной.

Для построения тематической модели коллекции документов  $D$  необходимо:

- найти множество тем  $T$ ,
- наблюдаемое распределение слов по тематикам  $p(w|t)$  для всех  $t \in T$ ,
- распределения тематик  $p(t|d)$  для всех документов  $d \in D$ .

В терминах тематического моделирования услуги, оказанные пациенту в ходе медицинского обслуживания, соотносятся со словами модели (Таблица 1). Маршрут пациента, состоящий из услуг, аналогичен документу со словами. Таким образом, обнаруженные алгоритмом скрытые темы интерпретируются как шаблоны, или паттерны, клинических путей пациента.

Гипотеза о независимости элементов выборки («мешка слов») эквивалентна предположению, что порядок терминов в документах не важен для выявления тематики, то есть тематику документа можно узнать даже после произвольной перестановки терминов.

Таким образом, метод нечёткой кластеризации позволяет разделить входной поток для системы с элементами самоорганизации, когда последовательность обхода специалистов определяет сам пациент.

Таблица 2.

**Соотнесение понятий тематического моделирования с объектами моделирования**

Термины тематического моделирования	Моделируемые объекты системы
$D$ – коллекция документов	$L$ – лог событий организации
$T$ – множество тематик	$Z$ – множество паттернов путей
$W$ – словарь терминов	$A$ – перечень оказываемых услуг
$d \in D$ – документ	$s \in S$ – клинический путь пациента
$t \in T$ – тематика	$z \in Z$ – паттерн (типовой путь)
$w \in W$ – слово	$a \in A$ – медицинская услуга
$\phi_{wt}$ – распределение слов в теме $t$	$\phi_{az}$ – распределение услуг в каждом паттерне $z$
$\theta_{td}$ – распределение тем в документе $d$	$\theta_{zs}$ – вероятности паттернов в каждом маршруте $s$

Предполагается, что появление услуг  $a$  в маршруте  $s$ , относящихся к паттерну  $z$ , описывается общим для лога событий  $L$  организации распределением  $p(a|z)$  и не зависит от документа самого маршрута. В терминах тематического моделирования это предположение называется гипотезой условной независимости:

$$p(a|s, z) = p(a|z); \quad (3)$$

$$p(s|a, z) = p(s|z); \quad (4)$$

$$p(s, a|z) = p(s|z)p(a|z) \quad (5)$$

Согласно определению условной вероятности, формуле полной вероятности и гипотезе условной независимости вероятностная тематическая модель:

$$P(a|s) = \sum_{t \in T} p(z|s)p(a|z) = \sum_z \phi_{az} \theta_{zs}, \quad (6)$$

где  $\phi_{az} = P(a|z)$  – вероятности активностей в каждом паттерне. Каждый маршрут задан  $s$  условным распределением на множестве всех паттернов  $z$ :

$\theta_{zs} = P(z|s)$  – вероятности паттернов в каждом маршруте. Эти распределения позволяют вычислить вероятность возникновения отдельной клинической активности у пациента

Подход АРТМ основан на идее многокритериальной регуляризации: позволяет строить модели, удовлетворяющие сразу нескольким ограничениям одновременно. Каждое ограничение формализуется в виде регуляризатора — оптимизационного критерия  $R_i(\Phi, \Theta) \rightarrow \max$ , зависящего от параметров модели.

Взвешенная сумма всех таких критериев:

$$R_i(\Phi, \Theta) = \sum_{i=1}^k \tau_i R_i(\Phi, \Theta) \quad (7)$$

Задача максимизации совместно с основным критерием правдоподобия:

$$L(\Phi, \Theta) = \sum_{s \in S} \sum_{a \in A} n_{sa} \log \sum_{z \in Z} \phi_{az} \theta_{zs} + R(\Phi, \Theta) \rightarrow \max_{\Phi, \Theta} \quad (8)$$

$$\sum_{a \in A} \phi_{az} = 1; \phi_{az} \geq 0; \quad (9)$$

$$\sum_{a \in A} \theta_{zs} = 1; \theta_{zs} \geq 0; \quad (10)$$

Для решения задачи регуляризованного правдоподобия используется EM-алгоритм с модифицированными формулами M-шага:

$$\phi_{az} \propto (n_{az} + \phi_{az} \frac{\partial R}{\partial \phi_{az}}) + \quad (11)$$

$$\phi_{az} \propto (n_{az} + \phi_{az} \frac{\partial R}{\partial \phi_{az}}) + \quad (12)$$

В диссертационном исследовании предложена стратегия регуляризации, основанная на оценке влияния каждого регуляризатора и их комбинации на качество итоговой модели. В процессе оптимизации на каждом этапе регуляризаторов добавляются последовательно, чтобы оценить вклад каждого, в отличие от рассмотренных ранее подходов [26- 29], при которых структура регуляризаторов во время оптимизации фиксирована.

Модифицирован стандартный EM-алгоритм для ARTM (Algorithm 2) за счёт добавления внешнего цикла, в котором выбирается наилучшая новая модель по метрике перплексии. Адаптация этой стандартной стратегии требует учёта особенностей данных. Данные предметной области с точки зрения регуляризации обладают большой размерностью, содержат большое количество числовых и категориальных признаков, тип объекта определяется на основании реализованного пути, представленного временным рядом.

Экспериментальным путём подобраны коэффициенты регуляризации, поскольку автоматическая коррекция стратегий регуляризации в ARTM пока является открытой научной проблемой.

Algorithm 1. EM- Алгоритм для ARTM (исходный)

---

**Вход:** коллекция документов  $D$ , число тем  $|T|$ ;  
**Выход:**  $\Phi, \theta$ ;

- 1 инициализировать вектор- столбцы  $\phi_t, \theta_d$  случайным образом;
- 2 **повторять**
- 3     обнулить  $n_{wt}, n_{td}$  для всех  $d \in D, w \in W, t \in T$ ;
- 4     **для всех**  $d \in D, w \in d$
- 5          $p(w|d) := \sum_{t \in T} \phi_{wt} \theta_{td}$ ;
- 6     **для всех**  $t \in T$
- 7          $p(t|d, w) := \phi_{wt} \theta_{td} / p(w|d)$ ;
- 8     увеличить  $n_{wt}, n_{td}$  на  $n_{dw} p(t|d, w)$ ;
- 9      $\phi_{wt} \propto (n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}})^+$  для всех  $w \in W, t \in T$ ;
- 10     $\theta_{td} \propto (n_{td} + \theta_{td} \frac{\partial R}{\partial \theta_{td}})^+$  для всех  $d \in D, t \in T$ ;
- 11 пока  $\theta, \Phi$  не сойдутся;

---

Algorithm 2. EM- Алгоритм для ARTM (модифицированный)

---

**Вход:** коллекция документов  $D$ , число тем  $|T|$ ;  
**Выход:**  $\Phi, \theta$ ;

- 1 инициализировать вектор- столбцы  $\phi_t, \theta_d$  случайным образом;
- 2 **повторять**
- 3     **для всех**  $r \in R'$

---

- 4                   обнулить  $n_{wt}, n_{td}$  для всех  $d \in D, w \in W, t \in T$ ;  
5                   для всех  $d \in D, w \in d$   
6                    $p(w|d) := \sum_{t \in T} \phi_{wt} \theta_{td}$ ;  
7                   для всех  $t \in T$   
8                    $p(t|d, w) := \phi_{wt} \theta_{td} / p(w|d)$ ;  
9                   увеличить  $n_{wt}, n_{td}$  на  $n_{dw} p(t|d, w)$ ;  
10                   $\phi_{wt} \propto (n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}})^+$  для всех  $w \in W, t \in T$ ;  
11                   $\theta_{td} \propto (n_{td} + \theta_{td} \frac{\partial R}{\partial \theta_{td}})^+$  для всех  $d \in D, t \in T$ ;

12                   $\phi'_{wt}, \theta'_{td}$ - наилучшие

13   пока  $\theta, \Phi$  не сойдутся;

В результате серии экспериментов подобрана комбинация следующих регуляторов по равномерной сетке значений с эмпирически обоснованными коэффициентами:

- Декоррелирование распределений услуг в маршрутах;
- Сглаживание/разреживание распределений услуг в паттернах маршрутов;
- Сглаживание/разреживание распределений паттернов в маршрутах.

Применение этого более гибкого подхода для кластеризации клинических путей пациентов в исследованиях ранее не рассматривалось. Использовалась библиотека с открытым кодом BigARTM в Python, в основе которой лежит аддитивная регуляризация. Данные были преобразованы в Vowpal Wabbit формат, который принимает входные данные в определенной структуре:

метка | A feature1: значение1 | B feature2: значение2.

Этот формат адаптирован для разделения на категории или модальности при обучении модели. Модель была создана и обучена по начальному количеству тем  $T = 300$ . На основе рассчитанных параметров перплексии 63,97 и коэффициентов разреженности  $\Theta=0,44$  и  $\Phi=0,42$  было выбрано оптимальное количество кластеров, равное девяти.

Каждому уникальному пациенту была присвоена вероятностная оценка принадлежности к определенному кластеру. Например, для одного из пациентов выборки распределение по шаблонам клинических путей выглядит следующим образом:  $P_1=0,017998157$ ,  $P_2=0,059349068$ ,  $P_4=0,5676379$ ,  $P_6=0,35303143$ . Соответственно, следующий шаг пациента с вероятностью около 57% будет соответствовать поведенческому паттерну 4 кластера.

Метод нечеткой кластеризации позволяет добавить иерархическое представление маршрутов пациентов, отображая ресурсы медицинских

учреждений. Выделенные кластеры будут отправной точкой для улучшения прогноза потока прикрепленного контингента, а также для формирования рекомендаций по ресурсному оснащению больниц при развитии сервисов.

**Третья глава** посвящена имитационной модели. В имитационную модель входные потоки включаются поочередно: сначала проводят серию экспериментов на базе статистических показателей моделируемой медицинской организации, далее, входной поток заявок представляют прогнозными значениями посещаемости. Для каждого типа пациентов отдельно строят прогноз посещаемости в будущие периоды на основе авторегрессионных и нейросетевых моделей прогнозирования временных рядов или последовательных значений посещаемости в определённое время.

Одним из популярных средств для прогнозирования посещаемости является модель авторегрессионного сезонного интегрированного скользящего среднего SARIMA, которая была выбрана в виду высокой точности, показанной в исследованиях для решения аналогичных задач.

Также была построена и оценена модель Хольта-Уинтерса, учитывающая экспоненциальный тренд и аддитивную сезонность. В качестве моделей глубинного обучения выбраны управляемый рекуррентный блок GRU и модель LSTM.

Возможным улучшением представленных алгоритмов является создание мета-алгоритма (рис. 3). Рассмотрим простейший мета-алгоритм на примере линейной регрессии. В математическом виде линейная регрессия для данного случая выглядит следующим образом:

$$y_t = w_1 y_t^{EXP} + w_2 y_t^{SARIMA} + w_3 y_t^{LSTM} + w_4 y_t^{GRU} \quad (14)$$

где  $y_t^{EXP}$  – прогноз на  $t$ -ый период модели HWES;  $y_t^{SARIMA}$  – прогноз на  $t$ -ый период модели SARIMA;  $y_t^{LSTM}$  – прогноз на  $t$ -ый период модели LSTM;  $y_t^{GRU}$  – прогноз на  $t$ -ый период модели GRU;  $w_1, w_2, w_3$  и  $w_4$  – веса моделей при прогнозе.

Другим примером мета-алгоритма является полносвязная нейронная сеть, оптимизированная с помощью библиотеки `hyperopt`. На вход такой нейронной сети поступают прогнозы моделей  $y_t^{EXP}, y_t^{SARIMA}, y_t^{LSTM}, y_t^{GRU}$  с  $n$  нейронами на первом слое, а на выходе определяется конечный прогноз  $y_t$ . Нейронная сеть обучается с помощью обратного распространения ошибки.

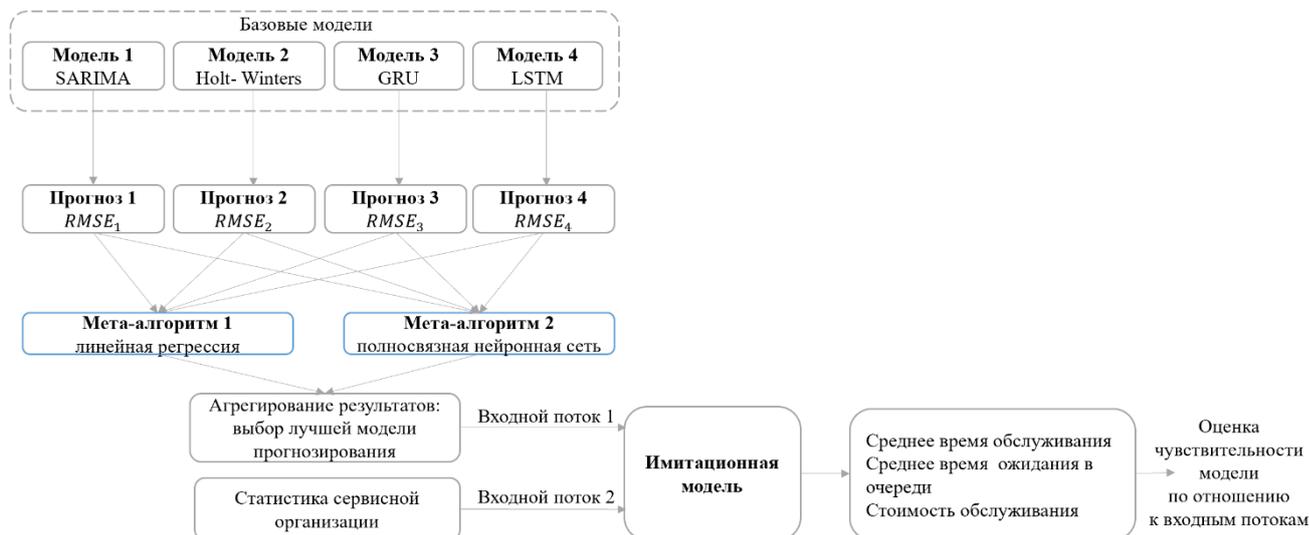
Для оценки точности прогнозирования выбран корень среднеквадратичной ошибки RMSE (Root Mean Squared Error). Данная

минимизирующая метрика (2) нередко применяется для исследования моделей, прогнозирующих потоки пациентов в медицинских учреждениях.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_i - O_i)^2} \quad (15)$$

(2)

где  $T_i$  – значения временного ряда;  $O_i$  – полученные прогнозные значения.



**Рисунок 4. Интеграция прогнозируемого входного потока в имитационную модель**

После оценки точности прогнозирования в имитационную модель интегрируются результаты мета-алгоритма с лучшей метрикой RMSE. На основе серии экспериментов оценивается чувствительность модели по отношению к прогнозируемому и фактическому входным потокам.

В качестве ограничений при выборе структуры сервисной организации в предлагаемой схеме моделирования используются бюджет, нормативный коэффициент загрузки персонала и время ожидания в очереди. Выбранная организационная структура должна обеспечивать минимизацию затрат времени на ожидание обслуживания и стоимости обслуживания. При этом также необходимо обеспечить минимизацию затрат на собственное функционирование. По результатам имитационного моделирования организация может сформировать обоснованную стратегию развития и обозначить дальнейшие этапы её совершенствования.

Прогнозные модели были обучены в облачной платформе Google Colaboratory и оптимизированы с помощью библиотеки hyperopt. В табл. 1 представлены результаты метрики RMSE по каждой модели и каждой группе. Наилучшие результаты наблюдаются у модели LSTM (выделены в таблице) по группам  $g_2, g_4, g_5, g_6$ , однако в некоторых случаях SARIMA и GRU

показывают лучшие показатели по группам  $g_1$  и  $g_3$  соответственно. Наилучшие показатели у LSTM-модели обусловлены особенностью архитектуры этой разновидности рекуррентных нейронных сетей, способной к фильтрации входной информации с помощью трёх типов структур, называемых воротами. Структура LSTM-модель способствует запоминанию информации на долгосрочные периоды и использованию её в прогнозе посещаемости пациентов рассмотренных групп.

Таблица 3.

**Результаты прогнозирования посещаемости в медицинской организации**

Модель	Результат					
<b>SARIMA</b>	4,6180	7,6787	<u>4,7832</u>	4,5814	5,0447	4,6760
<b>HWES</b>	4,6507	7,6230	4,9137	4,5863	5,0551	4,6828
<b>GRU</b>	4,5376	7,7650	4,8439	4,4483	5,0295	4,6324
<b>LSTM</b>	4,5380	<u>7,6721</u>	4,8566	<u>4,4444</u>	<u>4,9595</u>	<u>4,6205</u>

Таблица 4.

**Оценка мета-моделей для прогнозирования посещаемости в медицинской организации на основе метрики RMSE**

Модель	Результат					
<b>Linear regression</b>	4,5864	8,0750	4,7243	4,2856	4,8511	4,7925
<b>FCNN</b>	4,5370	7,3105	4,6550	4,3562	4,7629	4,7325

Проведённая серия экспериментов показала, что создание мета-алгоритма на основе полученных результатов базовых моделей SARIMA, Хольта-Уинтерса, управляемого рекуррентного блока GRU и LSTM значительно улучшает качество прогнозирования практически во всех группах пациентов.

**Заключение**

В предложенной диссертации исследуются проблемы определения рациональной ресурсной структуры сервисной организации, обеспечивающей принятие управленческих решений путем разработки прогнозных моделей для удовлетворения спроса на услуги при имеющихся экономических

ограничениях. Диссертация вносит важный вклад в повышение эффективности процесса обслуживания в медицинских организациях.

В-первых, проведенный анализ существующих подходов и методов к моделированию и совершенствованию рациональной организационной структуры сервисных организаций с учетом особенностей входных потоков субъектов обслуживания показал необходимость разработки альтернативного метода, основанное на данных организаций.

Во-вторых, предложен и разработан оригинальный метод анализа и типизации входного потока заявок на основе нечеткой кластеризации в рамках теории аддитивной регуляризации тематических моделей.

В-третьих, предложен адаптивный метод прогнозирования входных потоков, обеспечивающий учет влияния внешних факторов на входные потоки ЛПУ.

В-четвертых, приложена модернизированная модель обслуживания в ЛПУ на основе интеграции имитационной модели с разработанными методами анализа, типизации и прогнозирования входного потока заявок.

Результаты исследования позволили значительно улучшить точность моделирования для выбора рациональной структуры организации. Данная работа показывает возможность реализации предложенных решений в городских медицинских организациях.

## Список литературы

1. Осипов, Г. С. Исследование систем массового обслуживания с ожиданием в Anylogic/ Г. С. Осипов // Бюллетень науки и практики. – 2016. – № 10(11). – С. 139-151. – DOI 10.5281/zenodo.161072.
2. Пирогова И. Н. Теория очередей/И. Н. Пирогова, П. П. Скачков, Е. Г. Филиппова. – 2017.
3. Прокофьева Е.С., Мальцева С.В., Цю-жен-цин Д.А., Автоматизация управления ресурсами организации на основе имитационного моделирования с использованием прогнозирования входных потоков//Автоматизация. Современные технологии. 2021. Т. 75 № 6. С.272-280.
4. Прокофьева Е.С., Зайцев Р.Д. Анализ клинических путей пациентов в медицинских учреждениях на основе методов жесткой и нечеткой кластеризации // Бизнес-информатика. 2020. Т. 14. № 1. С. 19–31.
5. Воронцов К. В. Потапенко А. А. Регуляризация вероятностных тематических моделей для повышения интерпретируемости и определения числа тем // Компьютерная лингвистика и

интеллектуальные технологии: По материалам ежегодной  
Международной конференции «Диалог» (Бекасово, 4–8 июня 2014 г.)  
Вып.13 (20). М: Изд-во РГГУ, 2014. С.676–687.