Обучение с подкреплением для оптимизации распределения пациентов внутри больницы при чрезвычайных ситуациях

Глеб О. Бондаренко
Национальный исследовательский университет ИТМО
Olakola9@gmail.com

Анастасия Р. Попова
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
renoridoru@gmail.com

Алёна С. Сырых
Национальный исследовательский университет ИТМО
alyoshca.syryh@mail.ru

Егор В. Патока
Национальный исследовательский университет ИТМО
yegor-patoka@mail.ru

Евгений И. Гейченко
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
geychenko.1995@mail.ru

Владислав С. Павлюк
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
vlad.pavluk03.@mail.ru

Аннотация. **Статья посвящена разработке и оптимизации процесса обучения с подкреплением для повышения эффективности распределения пациентов в больницах в условиях экстренной медицины. В работе представлена методология, позволяющая формировать процесс обучения таким образом, чтобы обеспечивать адаптивное и оперативное принятие решений в условиях динамически изменяющейся обстановки. Анализируется влияние используемых методов на точность и эффективность распределения медицинских ресурсов. Экспериментальные результаты подтверждают практическую значимость предложенного подхода, демонстрируя его способность повышать скорость и качество оказания экстренной медицинской помощи.**

Ключевые слова: обучение с подкреплением, распределение пациентов, неотложная медицина, оптимизация, чрезвычайные ситуации

I. Введение

В последние годы интеграция обучения с подкреплением (RL) в медицинскую логистику продемонстрировала значительный потенциал в оптимизации потоков пациентов и распределения ресурсов. Традиционные модели, такие как дискретно-событийное моделирование и статические методы оптимизации, показали ограничения в динамической адаптации к изменениям в реальном времени [1] в чрезвычайных медицинских ситуациях. Эти ограничения включают отсутствие гибкости в реагировании на внезапные колебания в прибытии пациентов, доступности ресурсов [2] и ограничениях приоритетов. В экстренной медицине, где решения, принимаемые с учетом времени, имеют решающее значение, жесткие модели часто не могут учесть непредвиденные изменения в спросе на пациентов и пропускной способности больницы.

Существующие подходы к маршрутизации пациентов в основном опираются на принятие решений на основе правил и эвристические алгоритмы, которые, хотя и эффективны в структурированных средах, не позволяют адекватно учитывать непредсказуемость чрезвычайных ситуаций. Эти модели часто требуют длительной ручной калибровки и не способны обучаться на основе новых данных или динамически корректировать стратегии. Более того, современные имитационные модели, такие как сети Петри и дискретно-событийное моделирование, дают ценные сведения, но не позволяют адаптироваться в реальном времени, что крайне важно в сценариях реагирования на чрезвычайные ситуации.

Обучение с подкреплением предлагает многообещающую альтернативу, представляя адаптивные механизмы принятия решений, которые постоянно совершенствуются благодаря взаимодействию с окружающей средой. В отличие от традиционных методов оптимизации, агенты RL обучаются политике, которая максимизирует долгосрочное вознаграждение, что делает их хорошо подходящими для сложных, многоэтапных процессов принятия решений. Используя парадигму «состояние – действие – вознаграждение», RL может способствовать интеллектуальному распределению пациентов, динамически корректируя маршруты на основе данных реального времени, включая загруженность больницы, тяжесть состояния пациента и доступные медицинские ресурсы. Кроме того, модели на основе RL могут включать в себя методы обучения мультиагентов, что позволяет различным медицинским учреждениям сотрудничать в оптимизации регионального потока пациентов.

В данной работе представлена усовершенствованная модель маршрутизации пациентов на основе RL, разработанная для преодоления неэффективности традиционных методик. В предлагаемой системе интегрированы методы глубокого обучения с подкреплением для повышения адаптивности и быстроты реакции в динамично развивающихся средах здравоохранения. Благодаря использованию предиктивной аналитики, контуров обратной связи в реальном времени и стратегий совместного принятия решений, этот подход направлен на оптимизацию распределения медицинских ресурсов и минимизацию задержек пациентов, что в конечном итоге повышает общую эффективность реагирования на чрезвычайные ситуации.

II. Постановка задачи

В системах экстренной медицинской помощи управление потоками пациентов требует динамической адаптации к изменяющимся условиям, таким как колебания притока пациентов, доступность ресурсов и загруженность отделения. Традиционные модели, включая сети Петри и дискретно-событийные симуляции, не способны реагировать в реальном времени из-за своей статичности. Reinforcement Learning (RL) предлагает решение, позволяя агенту итеративно оптимизировать маршрутизацию пациентов на основе текущих условий в больнице и целевого маркера.

Пусть состояние s_t в момент времени T представляет собой текущую обстановку в больнице, включая длину очереди, наличие ресурсов и состояние пациента. Действие at соответствует выбору следующего маркера m_{t+1} для маршрутизации пациента. Задача состоит в минимизации общей функции затрат (1):

$$Q^{\pi}(s_t, a_t) = \mathbb{E}[r_t + \lambda Q^{\pi}(s', a')[s_t, a_t] \qquad (1)$$

где $Q^{\pi}(s_t, a_t)$ – оптимальное значение Q для выбора действия; r_t – функция вознаграждения, штрафующая за перегруженность и поощряющая эффективное распределение пациентов; (s', a') – представляет все возможные действия и состояния.

Агент RL обучается оптимальным правилам π, которая выбирает наилучшее решение по маршрутизации на каждом шаге, адаптируясь к динамике больницы в реальном времени.

III. Метод

В рамках предложенной модели маршрутизации пациентов в экстренных медицинских ситуациях ключевым этапом является организация выборки данных и разработка маршрута. Для эффективного управления потоками пациентов и ресурсов необходимо учитывать динамические изменения в системе, такие как доступность медицинских ресурсов, загруженность отделений и приоритетность пациентов. В данной работе предлагается метод структурирования выборки данных и организации маршрутов таким образом, чтобы минимизировать вычислительные затраты и обеспечить адаптивность системы. Сама выборка состоит из чистой симуляции GOAP, в которой фиксируются текущие переходы других агентов и процент успешного прибытия в промежуточную точку назначения.

Для создания выборки данных, представляющей возможные пути следования пациента, был разработан подход, основанный на предварительном определении всех возможных путей между маркерами (точками принятия решений). В контексте медицинской логистики маркеры могут представлять собой отделения больницы, такие как приемное отделение, операционная, палаты интенсивной терапии и другие ключевые точки.

Все возможные маркеры в системе идентифицированы и пронумерованы. Например, маркер 1 может обозначать приемное отделение, маркер 2 – операционную, маркер 3 – отделение интенсивной терапии и так далее. Для каждого маркера определяются все возможные пути, включая прямое и обратное направления. Например, путь от маркера 1 к маркеру 2 (1→2) и обратный путь от маркера 2 к маркеру 1 (2→1). Формируется список всех возможных маршрутов, включая прямое и обратное направления.

Каждый агент (пациент) назначается на определенный маршрут в зависимости от его текущего состояния и целевого маркера. Например, если пациент находится в маркере 1 и его нужно направить к маркеру 3, ему назначается маршрут 1→3. Такой подход минимизирует вычислительные затраты, поскольку маршруты заранее определены и системе не нужно постоянно пересчитывать возможные пути для каждого агента.

Однако в процессе реализации модели мы столкнулись с рядом проблем, которые потребовали дополнительных решений для повышения стабильности и точности системы [3–4]. Одной из ключевых проблем было неадекватное поведение модели при выборе подходящего случая, если количество агентов в текущем случае значительно отличалось от количества агентов в выборке. Это приводило к тому, что модель неверно интерпретировала данные, что, в свою очередь, снижало точность маршрутизации и адаптивность системы. Например, если в выборке преобладали случаи с большим числом агентов, а текущий случай содержал гораздо меньшее число агентов, модель могла ошибочно назначить ему неподходящий маршрут, что негативно сказалось на общей производительности системы.

Другая проблема заключалась в поведении системы, когда новый случай настолько сильно отличался от данных в выборке, что модель начинала вести себя непредсказуемо. Это происходило потому, что система не была достаточно устойчива к выбросам и не имела механизмов для обработки случаев, значительно отличающихся от типичных сценариев. Например, если новый случай содержал координаты агента, выходящие за пределы типичного диапазона, модель могла либо назначить ему неправильный маршрут, либо вообще не обработать его, что приводило к сбоям в работе системы.

Для решения этих проблем было предложено использовать кластеризацию данных перед применением RL-модели. Такой подход позволяет проанализировать текущую ситуацию и определить, существует ли подходящий случай в выборке, прежде чем направить агента к следующему маркеру. Кластеризация на основе K-средних объединяет данные в кластеры, что позволяет системе работать с облаками точек, представляющими

типичные сценарии, а не отдельные случаи. Это решает проблему разницы в количестве агентов, поскольку система теперь ориентируется не на точное совпадение, а на принадлежность к определенному кластеру, что делает ее более гибкой и устойчивой к небольшим отклонениям.

Для работы с удаленными случаями был введен порог отклонения, который определяет, насколько новый случай может отличаться от данных в кластере. Если расстояние от нового случая до ближайшего кластера превышает порог, система классифицирует его как «неподходящий» и исключает из обработки. Этот механизм предотвращает непредсказуемое поведение модели и повышает ее надежность.

Для демонстрации работы системы было проведено моделирование, в котором рассматривался случай, существенно отличающийся от существующих как по количеству агентов, так и по разбросу их координат по маркерам. Целью симуляции было показать, как система правильно идентифицирует и отвергает такие случаи, используя введенные механизмы кластеризации и порога отбраковки.

В рамках моделирования был создан новый случай, в котором координаты агентов выходили за пределы типичного диапазона, а их количество значительно отличалось от данных в выборке. Этот случай был обработан системой, которая, используя метод кластеризации K-means, определила, что новый случай не принадлежит ни к одному из существующих кластеров. На основании заданного порога отклонения система классифицировала его как «неприемлемый» и исключила из дальнейшей обработки.

Для наглядности результаты моделирования были визуализированы с помощью PCA, который уменьшил размерность данных и представил их в двумерном пространстве. На рис. 1 показаны кластеры, центры кластеров и новый случай, который был отклонен системой. Кластеры обозначены разными цветами, центры кластеров – черными крестиками, а отклоненный случай – красной звездочкой. Это визуальное представление демонстрирует, как система правильно определяет и обрабатывает случаи, которые значительно отличаются от данных в выборке.

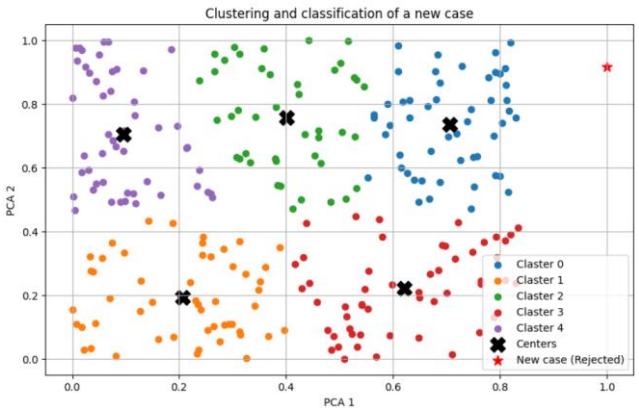

Рис. 1. Демонстрация моделирования кластеризации

IV. Результаты

В результате экспериментов и моделирования было выявлено, что использование технологии Reinforcement Learning в чистом виде без дополнительных механизмов анализа данных не всегда демонстрирует достаточную эффективность (рис. 2). Основная проблема заключалась в том, что система могла вести себя непредсказуемо [5] в ситуациях, существенно отличающихся от данных в выборке. Например, агент мог преградить путь критически важному пациенту или выбрать обходной маршрут без уважительной причины, что приводило к неоптимальному распределению ресурсов и задержкам в оказании помощи.

Эти проблемы были связаны с тем, что модель RL, обученная на определенных сценариях, не всегда могла правильно адаптироваться к новым, незнакомым ситуациям. В случаях, когда текущая ситуация сильно отличалась от данных в выборке, система могла принимать неверные решения, что снижало ее надежность и эффективность [6].

Для решения этой проблемы был предложен гибридный подход, сочетающий RL с кластеризацией и GOAP. Кластеризация позволяет системе проанализировать текущую ситуацию и определить, есть ли в выборке достаточно похожий случай. Если аналогичного случая нет, система автоматически переключается на GOAP, который строит маршрут на основе стандартных правил и логики. Как только агент достигает следующего маркера, система снова анализирует ситуацию и, если подходящий случай есть, возвращается к использованию RL.

Такой подход решает проблему непредсказуемого поведения в незнакомых ситуациях. Когда система понимает, что текущий случай значительно отличается от данных в выборке, она переключается на GOAP, что позволяет избежать ошибочных решений. В то же время, как только ситуация становится достаточно похожей на данные в выборке, система снова использует RL, что обеспечивает адаптивность и оптимизацию маршрута.

Таким образом, сочетание RL, кластеризации и GOAP (рис. 3) позволяет системе эффективно работать как в типичных, так и в нестандартных ситуациях. Это делает модель более прочной, надежной и адаптивной, что особенно важно в динамично меняющейся медицинской среде. Результаты моделирования подтвердили, что такой подход значительно повышает эффективность системы и минимизирует риски непредсказуемого поведения.

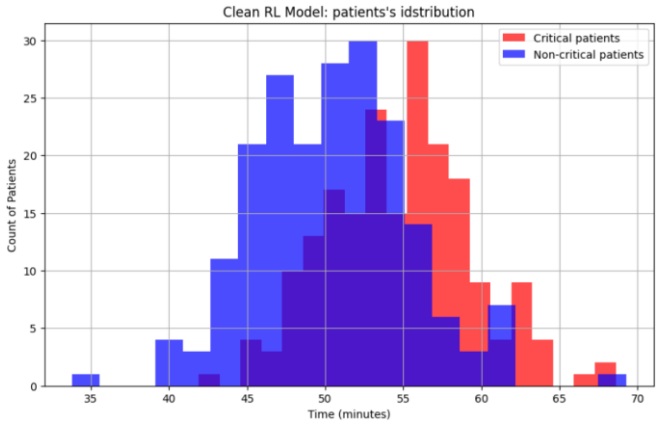

Рис. 2. Демонстрация работы чистой модели RL

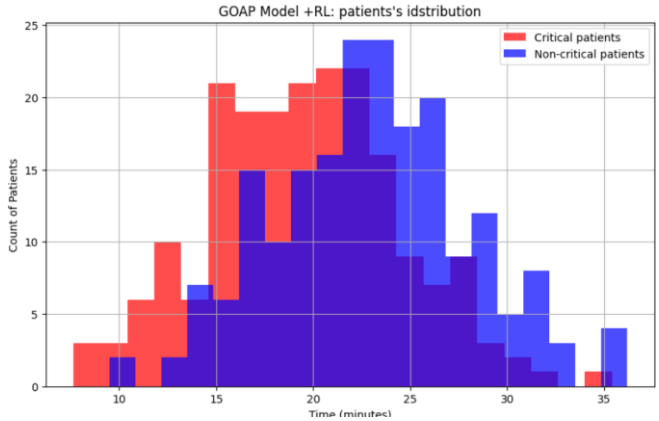

Рис. 3. Демонстрация работы модели GOAP+RL

V. Выводы

В данной работе был предложен и реализован гибридный подход к управлению маршрутизацией пациентов в экстренных медицинских ситуациях, сочетающий методы Reinforcement Learning, кластеризации и GOAP. Основной целью разработки было создание системы, способной эффективно адаптироваться к динамически изменяющимся условиям, таким как колебания числа пациентов, доступность ресурсов и загруженность отделений, при этом минимизируя риски непредсказуемого поведения.

Чистое использование RL, несмотря на его потенциал в адаптивном принятии решений, показало свою ограниченность в ситуациях, значительно отличающихся от данных в выборке. Это приводило к принятию неоптимальных решений, таких как блокирование путей для критически важных пациентов или выбор неоправданно длинных маршрутов. Для устранения этих проблем был реализован механизм кластеризации, который позволяет системе анализировать текущий случай и определять, насколько он соответствует имеющимся данным. В тех случаях, когда текущий случай оказывается слишком далеким от данных в выборке, система автоматически переключается на GOAP, обеспечивая стабильность и предсказуемость.

Ключевым преимуществом предложенного подхода является его гибкость. Система способна динамически переключаться между RL и GOAP в зависимости от текущей ситуации, что позволяет ей эффективно работать как в типичных, так и в нестандартных условиях. Это особенно важно в медицинской среде, где ошибки маршрутизации могут иметь серьезные последствия.

Результаты моделирования подтвердили эффективность гибридного подхода. Система успешно выявляла и отбрасывала случаи, которые существенно отличались от данных в выборке, а также корректно переключалась между RL и GOAP, обеспечивая оптимальную маршрутизацию. Это делает модель более устойчивой, надежной и адаптивной, что способствует повышению общего качества медицинской помощи.

В будущем планируется расширить функциональность системы, включив в нее дополнительные параметры, такие как приоритетность пациентов и доступность специализированных ресурсов. Кроме того, предполагается оптимизировать алгоритмы кластеризации и RL для повышения производительности системы в условиях реального времени. Предложенный подход может быть адаптирован не только для медицинской логистики, но и для других областей, где требуется адаптивное управление в динамически меняющихся условиях.

Список литературы

[1] A.R. Popova, G.O. Bondarenko, A.S. Syryh, V.S. Pavluk, E.I. Geichenko and V.D. Burlaka, "Intelligent Algorithms for Gaming AI: Research Into Decision-Making Methods and Agent Behavior," 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE), Novosibirsk, Russian Federation, 2024, pp. 830-833, doi: DOI: 10.1109/PIERE62470.2024.10805043

[2] F. Uwano, "A Cooperative Learning Method for Multi-Agent System with Different Input Resolutions," 2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR), Batu Pahat, Malaysia, 2021, pp. 84-90, doi: 10.1109/ISAMSR53229.2021.9567835.

[3] Yogesh Hole, Alaa Shather, Ali Hussein Alrubayi, Hawraa Ali Sabah, Hayder Al-Ghanimi, Haider Alchilibi, "Incorporation of Agents with Ontology to Develop a Healthcare Assessment Provision System", 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), vol.6, pp.2724-2728, 2023.

[4] J. Tanda, A. Moustafa and T. Ito, "Cooperative Behavior by Multi-agent Reinforcement Learning with Abstractive Communication," 2019 IEEE International Conference on Agents (ICA), Jinan, China, 2019, pp. 8-13, doi: 10.1109/AGENTS.2019.8929151.

[5] W. Sause, "Coordinated Reinforcement Learning Agents in a Multi-agent Virtual Environment," 2013 12th International Conference on Machine Learning and Applications, Miami, FL, USA, 2013, pp. 227-230, doi: 10.1109/ICMLA.2013.46.

[6] Y. Liu, Y. Dou, S. Shen and P. Qiao, "Global-Localized Agent Graph Convolution for Multi-Agent Reinforcement Learning," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 3480-3484, doi: 10.1109/ICASSP39728.2021.9414993.