МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ 2022 год, том 34, номер 11, стр. 107-122

ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ ПАНДЕМИИ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

© *2022 г.* ***В.А. Судаков*[1], *Ю.П. Титов*[2]**

[1] Институт прикладной математики им. М.В. Келдыша РАН
[2] Российский экономический университет им. Г.В. Плеханова

sudakov@ws-dss.com

DOI: 10.20948/mm-2022-11-07

Предлагается подход, позволяющий на основе достаточно простых моделей предложить методику прогнозирования решения органов управления о количестве необходимых медицинских центров для борьбы с пандемией. В основе него лежит идея о том, что решение об открытии нового центра принимается не сразу при переполнении существующих центров, а с некоторым опозданием. Таким образом, правительство пытается минимизировать риски ненужного открытия и принимает такое решение, понимая, что перегруженность существующих центров не прекратится в ближайшее время. Это решение можно предсказать, обучив модель на исторических данных, полученных из открытых источников. Такая модель разработана нами. Она позволяет прогнозировать количество медицинских центров на основе прогноза количества госпитализированных пациентов за 14 дней. Предложены подходы к прогнозированию количества госпитализированных пациентов с точностью, достаточной для того, чтобы модель предсказывала количество медицинских центров. Модели протестированы на данных из открытых источников, полученных для Рязанской области. Для модели прогноза количества открытых медицинских центров в Рязанской области определены штрафные функции и рассчитаны соответствующие коэффициенты.

Ключевые слова: поддержка принятия решений, прогнозирование количества медицинских центров, управление ресурсами, штрафная функция.

PANDEMIC FORECASTING BY MACHINE LEARNING IN A DECISION SUPPORT PROBLEM

***V.A. Sudakov*[1,2], *Yu.P. Titov*[2]**

[1] Keldysh Institute of Applied Mathematics of Russian Academy of Sciences
[2] Plekhanov Russian University of Economics

The paper proposes an approach that allows, based on fairly simple models, to propose an approach to predicting the decision of the governing bodies on the number of necessary medical centers to combat the pandemic. This approach is based on the idea that the decision to open a new center is not made immediately with the overflow of existing centers, but with some delay. Thus, the government is trying to minimize the risks of unnecessary opening and makes this decision, realizing that the congestion of existing centers will not end in the short term. This decision can be predicted by training the model on historical data obtained from open sources. We have developed a model that can be trained on historical data and allows forecasting the number of medical centers based on a forecast of the number of hospitalized patients for 14 days. Approaches are proposed for predicting the number of hospitalized patients with accuracy sufficient for the model to predict the number of medical centers. The models were tested on data from open sources obtained for the Ryazan region. For the forecast model for the number of open medical centers in the Ryazan region, penalty functions are determined and the corresponding coefficients are calculated.

Keywords: decision support, predicting the number of medical centers, resource management, penalty function.

1. Введение

Пандемия SARS-Covid-19 несомненно до сих пор является серьезным вызовом не только для современной медицины, но и всей науки в целом. Появились центры ежегодного мониторинга ситуации. Данные центры не только собирают статистику, но и определяют ее точность. Эти данные широко используются как для «громких заголовков» и принятия решений, так и для работы научного сообщества с целью анализа и прогнозирования динамики развития ситуации.

В настоящий момент основным источником информации являются официальные сведения, поставляемые медицинскими организациями и правительственными органами [1-3]. В результате практически во всех агрегаторах информация будет близкая. Тем не менее отмечаются ошибки в статистической информации, поступающей от органов власти. Наиболее авторитетные агрегаторы проверяют получаемую информацию по разным источникам и пытаются оценить ее правдоподобность. К таким агрегаторам можно отнести Our World in Data [4,5] и Worldometers [6]. Эти мировые агрегаторы собирают статистику по общему числу заболевших, выздоровевших и умерших, количеству вакцинированных, протестированных и госпитализированных. При этом агрегация данных происходит по странам, а для выявления относительных показателей учитывается население данных стран.

При этом указывается, что статистика неточная и может различаться между странами и даже округами/штатами. Так Worldometers [6] пишет, что

такой параметр, как «Количество выздоровевших» является неточным, так как основываться на разных определениях. ВОЗ рекомендует следовать критериям: симптомы проходят + 2 отрицательных теста в течение 24 часов или симптомы исчезают + дополнительные 14 дней, но это только рекомендация. В некоторых странах, когда пациент выписывается из больницы, он считается «выздоровевшим», даже если тест не проводится. Некоторые чиновники здравоохранения теперь считают, что любой человек, у которого был диагностирован COVID-19 три или более недель назад и который не умер, выздоровел от этой болезни. Ввиду этого показатель «Активные случаи», который зависит от количества выздоровлений, может зависеть от этого врожденного недостатка подсчета количества выздоровевших как для многих стран, так и для всего мирового подсчета. Активные случаи = (общее количество случаев) - (общее количество смертей) - (выздоровевших). Эта цифра представляет собой текущее количество людей, у которых выявлено и подтверждено заражение вирусом. Она может увеличиваться или уменьшаться и является важным показателем для органов общественного здравоохранения и реагирования на чрезвычайные ситуации при оценке потребностей в госпитализации.

Our World in Data рассматривает также вопрос определения числа заболевших [4]. В эпидемиологии случаи COVID-19, как и другие заболевания, широко определяются в рамках трехуровневой системы:

1. Подозрением на болезнь считается тот случай, когда проявляются клинические признаки и симптомы COVID-19, но не было лабораторных исследований.

2. Вероятный случай. Подозрение на болезнь с эпидемиологической связью с подтвержденным случаем. Это означает, что у кого-то проявляются симптомы COVID-19 и он либо находился в тесном контакте с положительным пациентом, либо находится в зоне, особенно пострадавшей от COVID.

3. Подтвержденный случай – это «человек с лабораторным подтверждением инфекции COVID-19», как объясняет Всемирная организация здравоохранения (ВОЗ). Как правило, для подтверждения случая у человека должен быть положительный результат лабораторных анализов. Это верно независимо от того, проявились ли у них симптомы COVID-19 или нет.

Для оценки масштабов вспышки необходимо знать фактическое количество случаев заражения, но фактическое количество случаев зависит от количества проведенных тестов и увеличение объемов тестирования населения увеличит объемы подтвержденных случаев, и статистика заболеваний будет неуклонно расти. Данный подход к сбору статистики делает ее неод-

нородной. Частое тестирование медицинского персонала и редкое тестирование групп граждан, не обязанных проходить тестирование искажает статистику по заболеваемости. Но эти данные можно использовать для случаев госпитализации, так как при госпитализации больному практически всегда делают тестирование ПЦР.

Спорным является также и параметр смертности от COVID-19. Часто пациенты умирают от сопутствующих заболеваний при уже вылеченном заболевании COVID-19, иногда наоборот. Такая ситуация вызвана правилами диагностирования причины смерти, которые могу различаться. Кроме того, пандемические заболевания накладывает на медицинский персонал дополнительные требования, что приводит к политическим решениям относительно ведения медицинской карты пациента и выполняемых процедур.

В результате такой показатель, как «Активные случаи» достаточно неточно описывает текущее положение дел, и определять количество потенциальных больных только на основе этого показателя неверно. Но данная статистика, хотя и имеет достаточно высокий уровень погрешности, используется органами власти для принятия решений. Для задачи предсказания управленческих решений можно применять данную статистику, так как решения определяются не на одном точном показателе, а на динамике и комплексном состоянии.

Для Российской Федерации управленческие решения по открытию новых отделений в рамках имеющихся медицинских центров осуществляется на местах или в региональном руководстве. При этом если не считать специализированные центры по борьбе с пандемией, эти решения принимаются исходя из региональной статистики [2], а также статистики в рамках одного медицинского центра. Количество госпитализированных человек по округам можно оценить с помощью оценки отношения количества госпитализированных к количеству заболевших, вычисленных по всей России. За счет единых требований и правил госпитализации пациентов можно предположить, что данная оценка будет достаточно достоверной. Информацию по количеству госпитализированных пациентов возьмем с агрегатора Worldometers [6]. Полной информации по России там нет, но есть несколько стран (указаны ISO коды стран), указывающих объемы недельной госпитализации больных. Отношение количества госпитализированных к общему количеству новых случаев заболевания $weekly_hosp_admissions / \sum_{7 дней} new_{cases}$ параметры агрегатора Worldometers) (рис.1). Из рисунка видно, что значение показателя после октября 2020 г. остается стабильным (кроме всплеска в Израиле). Для данной статистики определим доверительные интервалы ма-

тематического ожидания процента госпитализированных при доверительной вероятности равной 0.95 (рис.2) На начальном этапе тестирование ПЦР проводилось в основном для подтверждения симптомов и диагноза, в основном в медицинских учреждениях, поэтому и процент госпитализированных высокий – более 15%. В дальнейшем тестирование стало необходимым правилом для множества действий, в некоторых странах оно стало всеобъемлющим. В результате стали выявляться заболевшие, которые не требовали госпитализации из-за отсутствия тяжелых симптомов. В результате процент госпитализированных сначала снизился до 7.5%, а позже упал ниже 5%-й отметки от общего числа заболевших. Данный показатель близок и к показателю официального агрегатора информации по России [1].

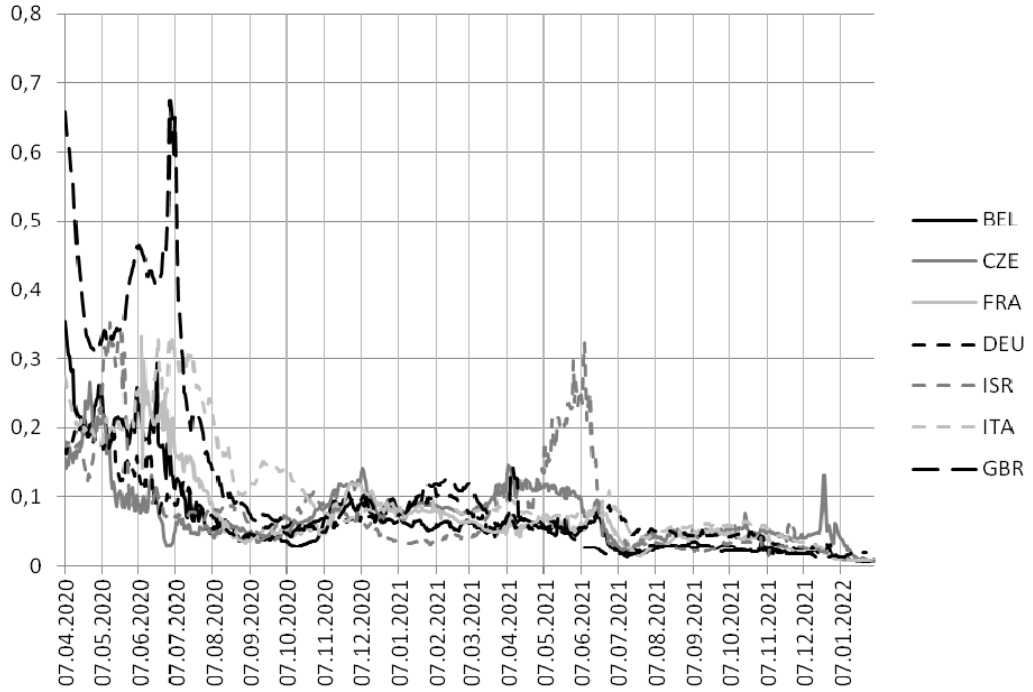

Рис.1. Отношение количества госпитализированных к общему количеству новых случаев заболевания для стран: Бельгия, Чехия, Франция, Германия, Израиль, Италия и Великобритания.

С помощью данных значений можно предположить объем госпитализированных по стране. Эти цифры помогают принимать административные решения на государственном уровне. Для анализа предполагаемого уровня требуемой госпитализации для отдельно взятого региона или медицинского учреждения необходимо взять статистику выявленных случаев заболеваний, взятую с агрегатора Our World in Data [5] по определенному региону, например, по Рязанской области, и определить объем госпитализации, используя параметр «Confirmed cases». Используя данный коэффициент, мож-

но оценить объем новых поступлений в медицинские учреждения. Но при этом данный показатель не определяет, сколько пациентов будет находиться в медицинских учреждениях.

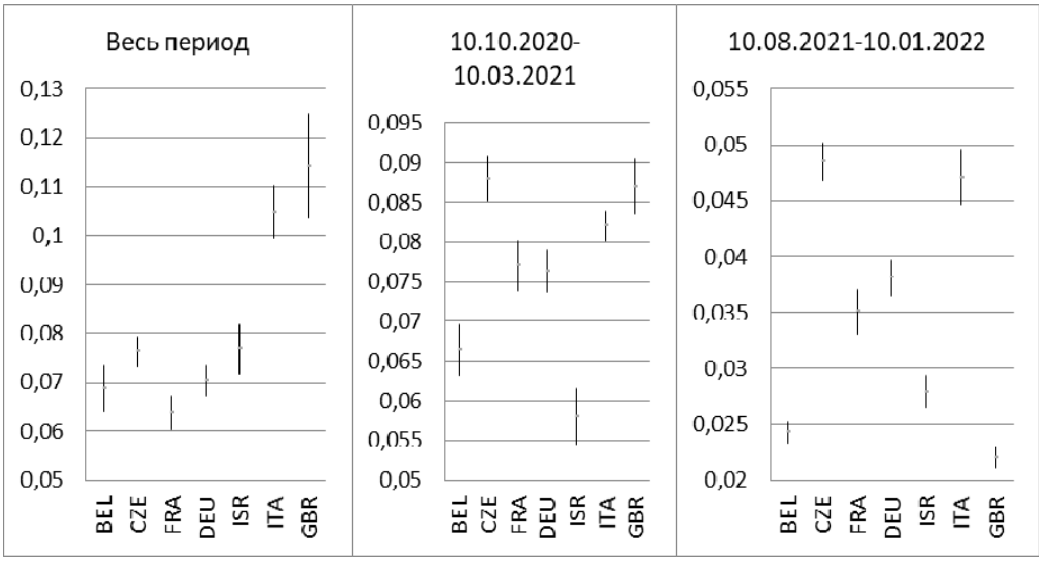

Рис.2. Среднее значение и доверительный интервал отношения количества госпитализированных к общему количеству новых случаев заболевания.

Агрегатор Our World in Data предоставляет информацию также о количестве «Активных случаев», т.е. общем количестве больных COVID-19 на определенную дату. Из официального сайта Правительства Рязанской области [2] можно получить информацию о количестве пациентов в медицинских учреждениях. На основании имеющейся информации можно вычислить коэффициент, определяющий процент больных, проходящих лечение в медицинских центрах, к числу заболевших COVID-19 (рис.3). Коэффициент не опускается ниже 10%, так как учитывается время пребывания в медицинском учреждении. В пиковом значении половина всех больных COVID-19 находились в медицинских учреждениях. Эти пики связаны с запаздыванием выписки из медицинского учреждения по сравнению с выздоровлением в домашних условиях.

На рис.4 приведены графики количества больных (штрих-пунктирная линия, правая ось) и госпитализированных (сплошная линия, левая ось). На сплошной линии отчетливо видны предельные возможности госпитализации. Это 220, 450, 660, 900 и более 1250 пациентов. Объем в 220 пациентов определяет комфортное количество пациентов для обслуживания медицинскими центрами. При снижении количества пациентов возможна дополнительная госпитализация в легких случаях заболевания. Объем в 450 определяет предельную нагрузку существующих отделений. Рост заболеваний в

мае-июне 2021 г. достиг предела (28-29 июня) в 450 госпитализированных. Далее потребовалось открытие нового центра, который стал быстро наполняться. В результате 3 июля количество госпитализированных увеличилось на 154 человека и через неделю достигло следующей отметки в 660 человек. Но количество заболевших увеличивалось, и нагрузка на существующие отделения не уменьшалась. В результате открываются еще отделения (16.10.2021 (900 пациентов) и 02.11.2021) и количество госпитализированных достигает 1250 пациентов.

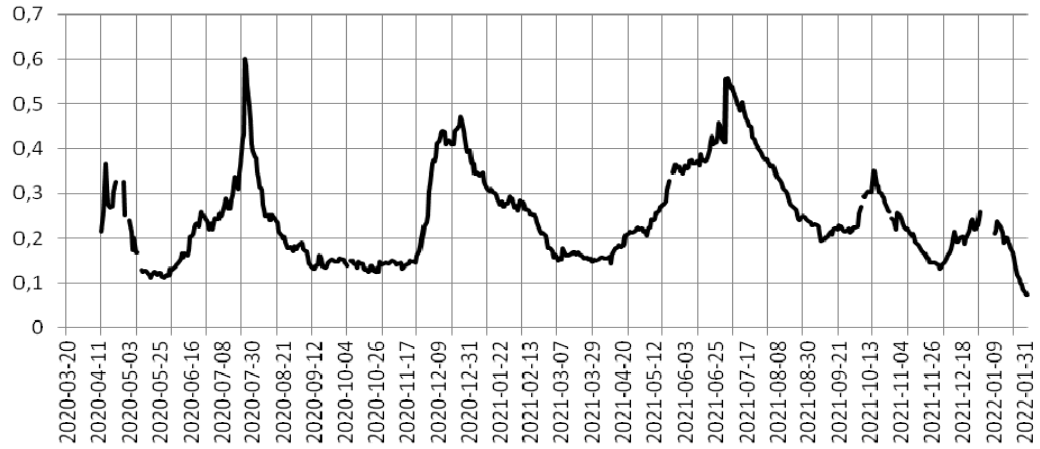

Рис.3. Отношение количества заболевших, находящихся на госпитализации, к общему числу заболевших в Рязанской области.

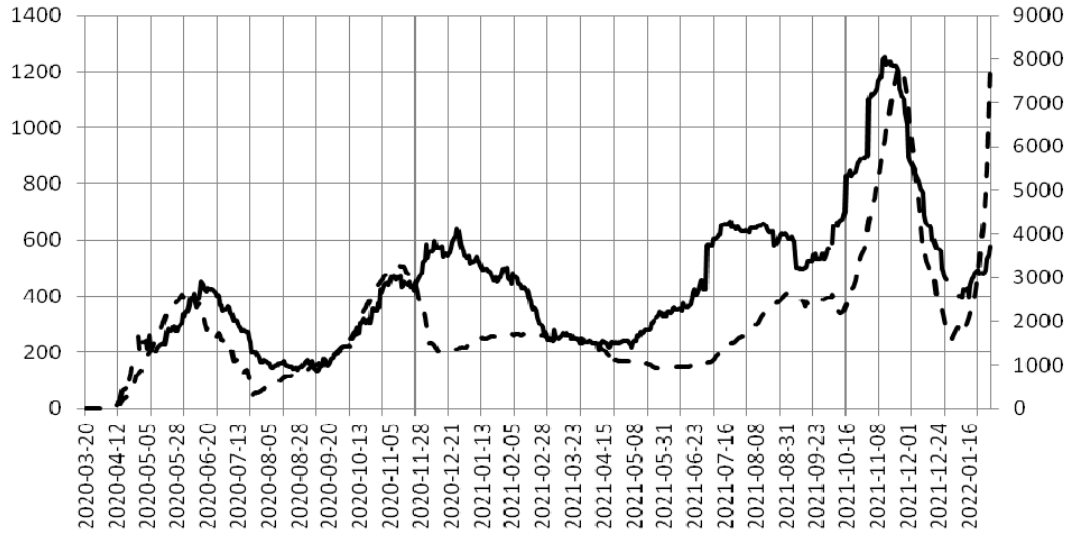

Рис.4. Количество заболевших и госпитализированных людей в Рязанской области.

В работе разрабатывается модель прогнозирования количества требуемых медицинских центров на срок более 14 дней. Количество медицинских центров можно описывать с помощью определенных центров, представленных в виде массива данных. Для всех работающих медицинских центров

можно описать коечный фонд, т.е. объемы пациентов, которые может принять данный медицинский центр. На основе полученных данных можно вычислить возможности области в госпитализации и сопоставить эти возможности с реальной и прогнозируемой обстановкой.

Так как фактически решается дискретная задача небольшого объема, то точность прогнозных значений может быть небольшой. Разрабатывается модель, которая определяет прогнозное значение открытых медицинских центров на основе прогноза количества госпитализированных пациентов. Прогнозное значение госпитализированных определяется на основе моделей прогноза и данных, полученных с открытых агрегаторов и официальных источников.

2. Методы прогнозирования применяемые для анализа пандемии

Современные работы, связанные с машинным обучением, нацелены в первую очередь на диагностику заболевания, чаще всего путем анализа снимков рентгенограмм [7-10]. В [11] предполагается применить многомерную линейную регрессию для прогнозирования количества случаев COVID-19 в течение четырнадцати дней. Вычисления проверяются с использованием точной организации, например, средней абсолютной ошибки (MAE), среднеквадратичной ошибки (RMSE) для наиболее затронутых регионов по всему миру. В [12] исторические данные были спрогнозированы с использованием экспоненциального сглаживания для выявления моделей сезонности и доверительных интервалов, окружающих каждое прогнозируемое значение. В [13] рассматриваются различные источники данных и возможность прогнозирования числа заболеваний исходя из различной информации. Для анализа часто применяют модель SEIRD [11,14-17]. Модель эпидемии SEIRD относится к классу так называемых компартментальных моделей, суть которых состоит в том, чтобы разделить популяцию на несколько групп (англ. Compartments), в нашем случае (англ. Susceptible) – восприимчивые, (англ. Exposed) – те, у кого болезнь находится в инкубационном периоде, (англ. Infectious) – больные, (англ. Recovered) – выздоровевшие, (англ. Dead) – умершие. Затем численность каждой из групп сопоставляется с переменной в системе дифференциальных уравнений, решая которую, можно спрогнозировать динамику развития эпидемии. Модификаций модели SEIRD достаточно много, например, SEIR – упрощенная модель, не учитывающая отдельно случаи выздоровления и смерти.

В работе [18] дается набор методов и моделей, применяемых для анализа и прогнозирования различных ситуаций, связанных с развитием COVID-19. В работе отмечаются методы анализа и прогнозирования, свя-

занные с большим набором данных, в том числе параметры атмосферы, информация в социальных сетях и другая. Тем не менее в работе отмечается положительная динамика при применении ограничительных мер. Отмечается работа Сяолинь Чжу и др. [19], в которой представлена пространственно-пандемическая модель для прогнозирования количества смертей. Это исследование направлено на создание модели прогнозирования, которая будет анализировать рост вируса в течение следующего месяца с учетом текущей динамики COVID-19. Но предложенные методы обычно имеют точность меньше, чем классические методы, основанные на статистике. Предлагаемые модели отлично прогнозируют на ближайшие дни, но прогноз на неделю или месяц уже не такой точный. При этом не рассматриваются вопросы прогнозирования объемов госпитализации с точки зрения работы медицинских учреждений. Для такой модели прогноз на пару дней будет мало полезен. Но предугадывание возможности открытия нового отделения в следующем месяце позволяет запланировать поставки и нанять необходимый персонал.

3. Описание модели

Обычно критериями качества моделей машинного обучения с учителем являются показатели, характеризующие точность прогноза. Но с точки зрения задач поддержки принятия решений результаты моделирования должны способствовать достижению критериев качества принимаемых решений. Конечно, хорошая точность прогноза должна обеспечивать построение хороших моделей принятия решений, но в условиях некоторых задач с малой или косвенной статистикой – например, прогнозирования потребления новых товаров, прогноз пандемий – хорошая точность прогноза недостижима и целесообразно переходить к критериям, характеризующим качество управленческих решений, получаемых на основе прогноза. Другая проблема широко распространенных точностных критериев – их симметричность. Однако во многих задачах положительная и отрицательная ошибка имеют различные последствия с точки зрения решений, принимаемых на основе прогнозов. Например, прогноз потребных запасов медикаментов: если прогноз завышен, то мы можем потратить излишние финансовые ресурсы, что безусловно не очень хорошо, но если прогноз занижен, то мы можем потерять человеческие жизни, что безусловно более критично.

Рассматривается модель, нацеленная на определение прогнозного значения открытых медицинских центров в определенный момент времени. Параметрами такой модели является множество медицинских центров (*MC*), расположенных в определенном округе или области. Каждый медицинский

центр обладает определенной вместимостью, не обязательно одинаковой C_i, $i \in (1..MC)$. Для удобства можно представлять множество медицинских центров как массив булевских значений. При этом X_i=1, $i \in (1..MC)$, если медицинский центр работает на прием пациентов с пандемией.

В рамках работы предлагается построить математическую модель для определения прогнозного значения работающих медицинских центров $\widetilde{X}$ на срок более двух недель с определенными возможностями госпитализации ($\widetilde{C_{sum}} = \sum_{i=1}^{MC} (\widetilde{X_i} C_i)$).

Прогноз предполагается строить на основе прогноза количества пациентов, находящихся в медицинских центрах $(\widetilde{P})$. Исходными данными такой модели могут служить исторические данные о количестве больных (*I*) и госпитализированных (*P*). Основными особенностями предлагаемой модели являются:

1. Имеется существующая инфраструктура, которая работает в штатном режиме и может обеспечить госпитализацию определенного объема $((MC_{shtat} \in MC,\ X_i = 1\ \ \forall i \in (1..MC_{shtat}))$.

2. Для медицинского учреждения есть объем госпитализации, связанный с «нормальным функционированием» (C_i), обычно данный объем меньше максимального возможного (часто ниже официально заявленной максимальной вместимости) в 2 раза ($C_{\max i}$). Тем самым практически сразу отделения наполняются до «нормального» объема и в случае роста числа заболевших происходит дальнейшее увеличение до максимального объема госпитализации.

3. Если открытие новых отделений происходит при достижении максимального уровня госпитализации, то новые отделения быстро наполняются пациентами.

4. Новые отделения открываются, если максимальный уровень пациентов держится определенное время. В июне 2020 г. первое достижение максимума (450 пациентов) продлилось 22 дня и не привело к открытию нового отделения. Второй пик с октября 2020 г. по февраль 2021 г. продлился 225 дней. Через 27 дней после достижения максимума в октябре 2020 г. был увеличен объем отделений, которые проработали до февраля. В июне-июле 2021 г. максимальный объем продержался 12 дней, после чего были открыты новые отделения.

Задачей, на которой осуществлено апробирование предложенного подхода, является анализ необходимого обеспечения отделений. Это касается закупки расходных материалов и найма персонала. Резкое открытие нового

отделения неэффективно с точки зрения закупки расходных материалов, не все материалы могут быть доставлены к открытию, могут быть куплены по завышенным ценам. Подобную задачу можно решать методами машинного обучения.

Исходными данными задачи являются:

1. Информация о медицинских центрах: $MC, C_i, C_{\max i}, MC_{shtat}$;

2. Информация о «исторических» данных работы центров $t \in (1,...,N)$, где N – объем исторических данных):

a) I_t – общее количество заболевших на момент времени *t*,

b) P_t – общее количество госпитализированных на момент времени *t*,

c) X_t – медицинские центры на момент времени *t*, задействованные в госпитализации пациентов.

В результате на каждый момент времени $t \in (N+1,...,N+32)$ необходимо определить прогнозное значение работающих медицинских центров ($\widetilde{X_t}$). Данный параметр определяется на основе прогнозного значения количества госпитализированных пациентов ($\widetilde{P_t}$) и машинного обучения. Так как расширения имеющихся медицинских центров является управленческим решением и зависит не только от объемов пациентов, то прогнозное количество медицинских центров будет определяться не только их возможностями, но и исторической информацией по скорости их открытия.

В рамках машинного обучения предлагается применять квантильную регрессию. Необходимо определить штрафные функции, с помощью которых можно оценить точность прогноза. Так как прогнозировать необходимо объемы поставок и количество работающих отделений медицинского центра, то штрафовать необходимо за недостаточность средств и медицинских центров для обслуживания всех случаев заболевания. При этом при «нормальном» функционировании отделений планирование новых поставок не требуется. Но при увеличении количества пациентов нагрузка на медицинские центры растет, при этом должна расти и штрафная функция. При достижении максимально возможного объема необходимо открытие нового отделения, если число пациентов будет расти. За открытие отделения без ресурсов необходимо установить максимальный штраф. Закупку материалов и подготовку персонала рекомендуется осуществлять за месяц до открытия отделения. Также необходимо рассмотреть возможность закупки небольшого количества расходных материалов и хранения их в работающих отделениях. Общий вид штрафной функции, вычисляемой на основе оценки теку-

щего количества госпитализированных ($\widetilde{P_t}$), нормальной ($\widetilde{C_{sumt}}$) и максимальной ($\widetilde{C_{\max t}}$) вместимости медицинских учреждений:

$$Q = \sum_{t=1}^{w1} w2_t (\widetilde{C_{sum_{N-t}}} - \widetilde{P_{N-t}})^{a_1}, \quad \text{если } \widetilde{C_{sum_{N-t}}} < \widetilde{P_{N-t}}, \tag{1}$$

$$Q = \sum_{t=1}^{w1} w2_t (\widetilde{P_{N-t}} - \widetilde{C_{sum_{N-t}}})^{a_2}, \quad \text{если } \widetilde{P_{N-t}} < \widetilde{C_{sum_{N-t}}} < \widetilde{C_{\max_{N-t}}}, \tag{2}$$

$$Q = \sum_{t=1}^{w1} w2_t (\widetilde{C_{\max_{N-t}}} - \widetilde{C_{sum_{N-t}}})^{a_2} (\widetilde{P_{N-t}} - \widetilde{C_{sum_{N-t}}})^{a_3}, \text{если } \widetilde{P_{N-t}} < \widetilde{C_{\max_{N-t}}}. \tag{3}$$

Здесь $\widetilde{P_{N-t}}$ – оценка количества госпитализированных пациентов, которая может быть получена как линейная комбинация прогноза и данных из официальных источников (если имеются). Прогнозное значение определяется исходя из общего количества случаев заболеваний и прогнозируемого количества случаев заболеваний с учетом коэффициента госпитализации. $\widetilde{X_{sum_{N-t}}}$ – определяет количество госпитализированных пациентов, которые могут принять медицинские учреждения при нормальной работе на *t*-й день ранее. Это значение определяется возможностями медицинских центров и вычисляется как сумма возможностей всех медицинских центров в районе, в том числе и вновь открытых. $\widetilde{C_{\max_{N-t}}}$ – определяет максимальную вместимость медицинских учреждений на *t*-й день ранее. Можно использовать оценку $\widetilde{C_{\max_{N-t}}} \approx 2\widetilde{X_{sum_{N-t}}}$; α – коэффициенты степени штрафной функции. Рекомендуется использовать α_1=2; α_2=1; α_3=2. Коэффициенты $w1$, $w2_t$ определяют конкретные зависимости для штрафной функции: $w1$ выражает «глубину» исторических данных, т.е. сколько предыдущих дней (измерений) учитывать при прогнозе. $w2_t$ – «значимость» исторических данных. Коэффициенты $w1, w2_t$ отличают отдельные практики работы правительственных органов по открытию новых отделений. Эти данные могут отличаться в различных регионах и поэтому штрафную функцию для каждого региона необходимо составлять отдельно. Для вычисления коэффициентов $w2_t$ можно воспользоваться машинным обучением, методом градиентного спуска. Для снижения размерности задачи (если анализировать месячную историю, то необходимо вычислять более 30 весов) можно объединить весовые коэффициенты $w2_t$ в блоки.

Обучение и тестирование модели предлагается проводить на основе имеющихся данных из официальных источников, используя официальные данные о количестве госпитализированных вместо прогнозных. По резуль-

татам оптимизации изменяется количество работающих медицинских центров $\widetilde{X_t}$ в момент времени t и параметры нормальной ($\widetilde{C_{sum_t}}$) и максимальной ($\widetilde{C_{\max_t}}$) вместимости медицинских центров таким образом, чтобы обеспечить требуемые объемы госпитализации.

4. Результаты работы модели

Оценку работы модели будем проводить на основе данных, полученных из официальных источников Рязанской области [2]. В Рязанской области расположен основной центр госпитализации при заболевании COVID-19 на 225 мест с максимальной загрузкой в 450. Второй центр позволяет увеличить количество мест до 350, а максимальное количество до 650. Еще два центра по 300 мест (максимальное значение) могут быть открыты в случае необходимости.

Модель обучалась на исходных данных для Рязанской области. В ходе обучения определялись значения весовых коэффициентов $w2_t$ на основе предложенных штрафных функций при $w1=30$. При этом определялись параметры нормальной и максимальной вместимостей всех открытых медицинских центров. На рис.5 линией с длинным штрихом отмечено прогнозное значение максимального объема возможной госпитализации на основе открытых медицинских центров при прогнозировании на 14 дней вперед. Из графика видно, что прогнозное значение объемов госпитализации обеспечивает реальные объемы.

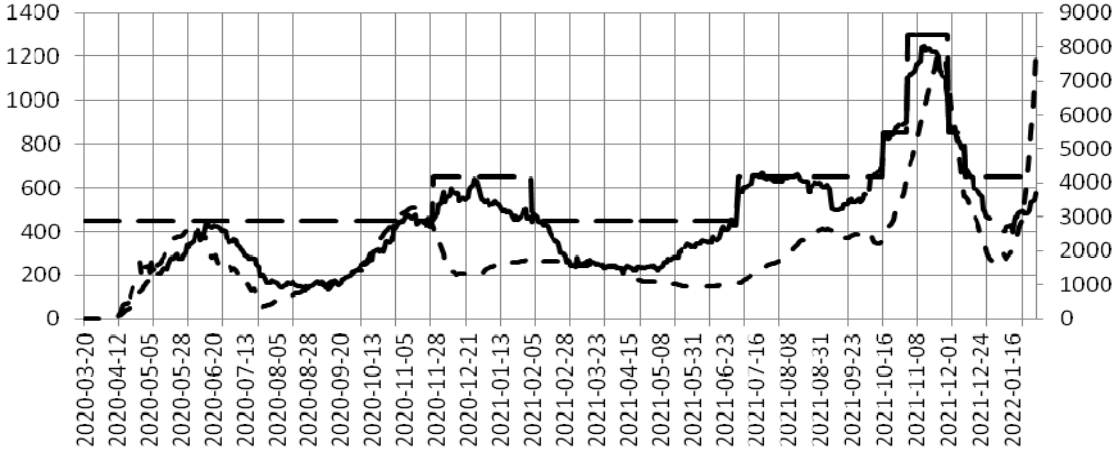

Рис.5. Количество заболевших, госпитализированных пациентов и прогноза объемов госпитализации медицинских центров в Рязанской области.

5. Прогнозирование объема госпитализации

Для использования модели необходимо знать количество госпитализированных. Прогнозирование данного параметра является очень сложной за-

дачей, так как достаточно точных моделей, описывающих время нахождения в медицинском центре при заболевании COVID-19 нет. Для работы таких моделей можно проводить оценку количества госпитализированных как некоторый процент от общего числа активных случаев, для прогнозирования которых имеются математические модели.

На основе данных из Our World in Data и официального сайта Рязанской области можно определить как объемы активных случаев, так и объемы госпитализации. Отношение этих величин, определяющее объемы госпитализированных больных в процентах, отображено на рис.3. Нормальный объем госпитализированных больных находится около отметки 20%. Пики определяют резкий рост числа находящихся на госпитализации пациентов по отношению к общему количеству больных и описывают вспышки пандемии. До апреля 2021 г. пики на графиках соответствуют уменьшению количества «Активных случаев». Длительность такого пика достигает 4-х месяцев. При этом можно сказать, что в пике около 50% заболевших находятся на лечении в медицинских центрах. По вычисленному прогнозному значению количества госпитализированных людей можно проводить прогноз работающих медицинских центров. Точность данной модели позволяет легко прогнозировать количество необходимых медицинских центров. Прогноз «Активных случаев» может быть получен из известных моделей, например, модель SEIRD.

Альтернативно можно использовать информацию по количеству госпитализированных пациентов каждый день. Эту информацию можно найти из множества источников, официальных сайтов [1], с агрегатора Worldometers и др. Этот параметр стационарен, как видно из рис.1 и 2. Для его использования необходимо построить модель прогнозирования времени госпитализации пациентов. В результате можно вычислить объемы госпитализированных пациентов на основе модели прогнозирования количества новых заболевших и оценки времени госпитализации.

6. Заключение

В работе предложен подход, позволяющий на основе достаточно простых моделей предложить подход к прогнозированию решения управляющих органах о количестве необходимых медицинских центров для борьбы с пандемией. Данный подход основан на идее о том, что решение об открытии нового центра принимается не сразу с переполнением существующих центров, а с некоторым запаздыванием. Таким образом, правительство старается минимизировать риски ненужного открытия и принимает это решение, понимая, что перегруженность существующих центров не закончится в

ближайшей перспективе. Данное решение можно спрогнозировать благодаря обучению модели на исторических данных, полученных из открытых источников. В основе модели лежит параметр, отвечающий за количество пациентов, находящихся на госпитализации. Предложены подходы, позволяющие спрогнозировать данный коэффициент с точностью, достаточной для работы модели прогнозирования количества медицинских центров. Так как управленческие решения зависят от установленных в правительстве норм и имеющихся ресурсов, то предполагается использование данной модели для оценки будущих параметров в том же районе, на котором и проводилось обучение.

СПИСОК ЛИТЕРАТУРЫ

1. Официальная информация о коронавирусе в России. Режим доступа: https://стопкоронавирус.рф/information/ (дата обращения 08.03.2022);
   Ofitsialnaia informatsiia o koronaviruse v Rossii. Available at: https://стопкоронавирус.рф/information/ (accessed 08.03.2022).
2. Правительство Рязанской области. Коронавирусная инфекция. Актуальная информация. Режим доступа: https://www.ryazangov.ru/governmentro/covid-19/ (дата обращения 08.03.2022);
   Pravitelstvo Riazanskoj oblasti. Koronavirusnaia infektsiia. Aktualnaia informatsiia. Available at: https://www.ryazangov.ru/governmentro/covid-19/ (accessed 08.03.2022).
3. Федеральная служба государственной статистики. Статистика против COVID-19. Режим доступа: https://rosstat.gov.ru/folder/81021 (дата обращения 08.03.2022);
   Federalnaia sluzhba gosudarstvennoj statistiki. Statistika protiv COVID-19. Available at: https://rosstat.gov.ru/folder/81021 (accessed 08.03.2022).
4. Our World in Data. Coronavirus Pandemic (COVID-19). https://ourworldindata.org/coronavirus (accessed 08.03.2022) [Our World in Data. Coronavirus Pandemic] (COVID-19). Available at: https://ourworldindata.org/coronavirus (accessed 08.03.2022).
5. Data on COVID-19 (coronavirus) by Our World in Data Режим доступа: https://github.com/owid/covid-19-data/tree/master/public/data/ (accessed 08.03.2022) [Data on COVID-19 (coronavirus) by Our World in Data] Available at: https://github.com/owid /covid-19-data/tree/master/public/data/ (accessed 08.03.2022).
6. Worldometer COVID-19 Data. https://www.worldometers.info/coronavirus/about/ (accessed 08.03.2022) [Worldometer COVID-19 Data] Available at: https://www.Worldometers.Info /coronavirus/about/ (accessed 08.03.2022)
7. *Mohapatra S.K., Assefa B.G., Belayneh G.* A SVM Based Model for COVID Detection Using CXR Image. In: Berihun M.L. (eds) Advances of Science and Technology. ICAST 2021. Lecture Notes of Institute for Computer Sci., Social Informatics and Telecommunications Eng., 2022, v.411. Springer, Cham. https://doi.org/10.1007/978-3-030-93709-6_24.
8. *Arowolo M.O., Ogundokun R.O., Misra S., Kadri A.F., Aduragba T.O.* Machine Learning Approach Using KPCA-SVMs for Predicting COVID-19. In: Garg L., Chakraborty C., Mahmoudi S., Sohmen V.S. (eds). Healthcare Informatics for Fighting COVID-19 and Fu-

ture Epidemics. EAI/Springer Innovations in Communication and Computing. 2022, Springer, Cham. https://doi.org/10.1007/978-3-030-72752-9_10.

9. *Assawab R., Elzaar A., El Allati A., Benaya N., Benyacoub B.* PCA SVM and Xgboost Algorithms for Covid-19 Recognition in Chest X-Ray Images. Advanced Technologies for Humanity. ICATH 2021. Lecture Notes on Data Engineering and Communications Technologies, 2022, v.110. Springer, Cham. https://doi.org/10.1007/978-3-030-94188-8_14.
10. *Sowmya Sundari L.K., Syed T. Ahmed, Anitha K., Pushpa M.K.* COVID-19 Outbreak Based Coronary Heart Diseases (CHD) Prediction Using SVM and Risk Factor Validation, Innovations in Power and Advanced Computing Technologies (i-PACT), 2021, p.1-5, https://doi.org/10.1109/i-PACT52855.2021.9696656.
11. *Nalini C., Shantha Kumari R., Bhuvaneswari M., Dheepthikaa V.S., Nandhini M.L.* Development of forecasting model for infectious disease (COVID-19), AIP Conference Proceedings 2021, p.2387, https://doi.org/10.1063/5.0069041.
12. *Saheed Oladele Amusat.* Forecasting the Epidemiological Impact of Coronavirus Disease (COVID-19): Pre-vaccination Era. medRxiv, 2021, No 03.17.21253791 https://doi.org/10.1101/2021.03.17.21253791
13. *Shinde G.R., Kalamkar A.B., Mahalle P.N. et al.* Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. SN COMPUT. SCI., 2020, №1, p.197. https://doi.org/10.1007/s42979-020-00209-9.
14. *Еремеева Н.И.* Построение модификации SEIRD-модели распространения эпидемии, учитывающей особенности COVID-19 // Вестник ТвГУ. Серия: Прикладная математика, 2020, выпуск 4, с.14-27, https://doi.org/10.26456/vtpmk602;
Eremeeva N.I. Postroenie modifikacii SEIRD-modeli rasprostraneniya epidemii, uchityvayushchej osobennosti COVID-19 // Vestnik TvGU. Seriya: Prikladnaya matematika, 2020, №4, p.14-27, https://doi.org/10.26456/vtpmk602.
15. *Rapolu T., Nutakki B., Sobha Rani T., Durga Bhavani S.* A Time-Dependent SEIRD Model for Forecasting the Transmission Dynamics in Infectious Diseases: COVID-19 / Proc. of Inter. Conf. on Data Science and Applications. Lecture Notes in Networks and Systems, 2022, v.287. Springer, Singapore. https://doi.org/10.1007/978-981-16-5348-3_33.
16. *Aliyeva T., Rzayeva U., Azizova R.* A SEIRD Model for Control of COVID-19: Case of Azerbaijan. SHS Web of Conf., 2021, №92, https://doi.org/10.1051/shsconf/20219201001
17. *Menda K., Laird L., Kochenderfer M.J. et al.* Explaining COVID-19 outbreaks with reactive SEIRD models. Sci Rep 2021, No 11. https://doi.org/10.1038/s41598-021-97260-0.
18. *Mahalle, Parikshit; Kalamkar, Asmita B.; Dey, Nilanjan; Chaki, Jyotismita; Hassanien, Aboul ella; Shinde, Gitanjali R.* Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. TechRxiv. Preprint, 2020. https://doi.org/10.36227/techrxiv.12101547.v1.
19. *Zhu X., Zhang A., Xu S., Jia P., Tan X., Tian J.* Spatially Explicit Modeling of 2019-n. CoV Epidemic Trend based on Mobile Phone Data in Mainland China, 2020.

Поступила в редакцию 20.04.22
После доработки 20.04.22
Принята к публикации 12.09.22