

Прогнозирование динамики эпидемии коронавируса (COVID-19) на основе метода прецедентов

В. В. Захаров, Ю. Е. Балыкина

Санкт-Петербургский государственный университет, Российская Федерация,
199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: Захаров В. В., Балыкина Ю. Е. Прогнозирование динамики эпидемии коронавируса (COVID-19) на основе метода прецедентов // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2020. Т. 16. Вып. 3. С. 249–259. <https://doi.org/10.21638/11701/spbu10.2020.303>

Описан метод прецедентов (Case-Based Reasoning) для прогнозирования будущих значений основных параметров эпидемии коронавируса в России, позволяющий строить краткосрочные прогнозы на основе аналогов динамики процентного прироста в других странах. Представлен новый эвристический метод оценивания продолжительности переходного процесса процентного прироста между заданными уровнями, учитывающий информацию о динамике эпидемиологических процессов в странах цепочки распространения. Разработан программный модуль Case-Based Rate Reasoning (CBRR) в среде MATLAB, реализующий предложенный подход и интеллектуальные авторские алгоритмы для построения траекторий прогнозируемых показателей эпидемии.

Ключевые слова: моделирование, прогнозирование, эпидемия COVID-19, процентный прирост, метод прецедентов, эвристика.

1. Введение. Особенность эпидемии нового вируса COVID-19 состоит в отсутствии статистики прошлых лет. В связи с этим возникает проблема адекватного использования имеющейся информации о параметрах развивающейся эпидемии, в том числе в других странах мира. Многие научные группы в США, Китае и Европе работают над созданием методов прогнозирования распространения нового вируса в краткосрочном периоде. 11 февраля 2020 г. Всемирная организация здравоохранения объявила о проведении Глобального исследовательского и инновационного форума и определила важнейшие цели исследований в области изучения распространения эпидемии [1]. В начале марта 2020 г. в журнале *Science Translation Medicine* вышла редакционная статья [2], в которой ее авторы сформулировали направления исследований, инновационные результаты которых должны способствовать превенции распространения эпидемии. Они также отметили, что всеобъемлющие математические модели, которые включают в себя сложные патогенные и общественно-значимые переменные, требуют значительного времени и усилий для их разработки и верификации (часто от месяцев до нескольких лет). Вместе с тем математические модели, прогнозирующие динамику регистрации новых случаев COVID-19 в режиме реального времени, начали появляться в журналах и на онлайн-ресурсах [3–8]. В статье [3] приводятся оценки масштабов эпидемии в Ухане и других городах, в том числе за пределами материкового Китая, в которые вирус мог быть экспортирован из Ухани. Авторы прогнозируют значения внутренних и глобальных рисков для здоровья населения от эпидемий на основе использования модели SEIR (Susceptible-Exposed-Infectious-Recovered), учитывая возможные сценарии профилактического вмешательства. В статье [5] изучена динамика распространения коронавируса в Индии с исполь-

зованием системы дифференциальных уравнений с постоянными коэффициентами и понятия базисного репродукционного числа, применяя для решения задачи оптимизации превентивных мер принцип максимума Понтрягина. В работе [6] обращено внимание на схожесть динамики общего числа инфицированных, выздоровевших и умерших людей в Китае и Италии. В ней также проведен анализ решений системы дифференциальных уравнений, принятой в модели SIRD (Susceptible-Infected-Recovered-Deaths), и отмечено, что, хотя модель SIRD является довольно грубой, ее использование дает хороший шанс отразить, по крайней мере, общие особенности эволюции эпидемии и спрогнозировать динамику в реальном времени. Предложенная методика направлена на прогнозирование наступления в Италии пика по приросту новых случаев инфицирования и количества смертей за весь период эпидемии. Авторы статьи [6] выдвигают гипотезу о том, что любая страна, которая становится театром эпидемического всплеска, может рассматриваться, по меньшей мере в первом приближении, как среда, различные группы населения которой взаимодействуют в соответствии с некоторыми общими правилами независимо от географических вариаций. Однако они демонстрируют, что генерируемые моделью SIRD траектории очень чувствительны по отношению к вариациям ее параметров, что значительно снижает качество прогнозирования, особенно на значительном горизонте времени. В [7] предложен пространственно-временной подход, основанный на использовании броуновского движения, и модель SBDiEM (Stereographic Brownian Diffusion Epidemiology Model) для прогнозирования динамики инфицирования, которые могут быть применены для создания системы мониторинга и противодействия пандемиям с помощью технологий искусственного интеллекта. Авторы статьи [8] используют количественную картину распространения заболевания COVID-19 в Китае в качестве тестового примера и данные об инфекции из восьми стран для оценки эволюции эпидемиологического процесса в каждой из этих стран. Такой подход основан на гауссовой гипотезе распространения вируса и базовой модели SIR (Susceptible-Infected-Recovered). Учитывая трудности в применении для прогнозирования динамики распространения COVID-19 детерминированных моделей типа SIR, SEIR, SIRD, которые построены на механизмах распространения вируса от индивидуума к индивидууму и используют оценки параметров распространения известных вирусов, наши усилия будут сосредоточены на поиске других методов. Важное наблюдение, сформулированное в [6], о схожести параметров эпидемии в Китае и Италии, а также анализ данных статистики распространения коронавируса в России, Италии, Испании, Франции и Великобритании привели нас к гипотезе о возможной зависимости скорости прироста основных показателей, регистрируемых в России, от аналогичных показателей в обследуемых странах. Поскольку эпидемиологические процессы в этих странах начались раньше, мы предположили, что, видимо, наиболее приемлем для анализа и прогнозирования метод прецедентов (Case-Based Reasoning), упомянутый в обзоре [9] и описанный в статье [10]. Более ранним вариантом такого метода является метод аналогий (Method of Analogues) [11]. Сравнивая часто применяемый на практике статистический подход и разработанный метод прецедентов, использующий не только статистику из прошлого, но и текущие данные, авторы [10] убедительно продемонстрировали преимущество предложенного ими метода при прогнозировании динамики эпидемии гриппа, волны которой характеризуются нерегулярными циклами, трудно предсказуемыми с помощью статистики эпидемий в прошлом.

Хронология исследования динамики эпидемии COVID-19 в России и прогнозирования ее развития в период с 1 мая по 30 июня 2020 г. отражена в п. 2. Математическая

модель прогнозирования процентного прироста общего числа выявленных заболевших рассмотрена в п. 3. В п. 4 приведено описание программного комплекса CBRR (Case-Based Rate Reasoning) для работы с базами данных по динамике пандемии в странах мира, формирования цепочек распространения (стран-предшественников и страны-последователя) и генерации прогнозных траекторий эпидемии на периоды продолжительностью 1–3 недели. В п. 5 обсуждаются полученные результаты, дискуссионные вопросы и задачи будущих исследований.

2. Эпидемия коронавируса в России. Моделирование и прогнозирование динамики. Процессу выбора модели и метода прогнозирования предшествовала исследовательская работа по мониторингу ситуации, после того как указом Президента РФ с 30 марта по 3 апреля 2020 г. были установлены нерабочие дни с сохранением за работниками заработной платы. Эта мера и последующие принятые решения сформировали ограничения для распространения эпидемии в России.

В первой аналитической заметке Центра интеллектуальной логистики, подготовленной 1 мая 2020 г. (<http://www.apmath.spbu.ru/cil/>), ставился вопрос: Каковы могут быть оценки даты достижения пика эпидемии в России?

В результате применения разработанного на основе метода прецедентов подхода был сделан следующий прогноз: выход на пик эпидемии по числу болеющих людей произойдет не раньше 16 мая. К этому времени общее количество инфицированных может преодолеть рубеж 160 000 человек. Фактически пик эпидемии по числу болеющих в России людей достигнут 25 мая, когда было зафиксировано 230 996 болеющих. В период с 5 мая по 15 июня 2020 г. были подготовлены восемь аналитических записок, в которых представлены прогнозы о будущей динамике развития эпидемии, учитывающие развитие эпидемии в Италии, Испании, Франции и Великобритании. Эти страны мы предложили называть «прецедентными», поскольку выбранные для анализа значения показателей достигались в них на 2–3 недели раньше, чем в России. Прогнозы и данные о подтверждении этих прогнозов внесены в таблицу.

3. Модель построения краткосрочных прогнозов на основе метода прецедентов. Официальная текущая статистика ежедневно публикует данные об общем количестве выявленных заболевших $I(t)$ (Infected) на текущий день t , выздоровевших от болезни $R(t)$ (Recovered), умерших $F(t)$ и количестве болеющих $A(t)$ (число активных случаев, Active Cases). Ключевыми параметрами, определяющими динамику данных статистики, являются значения процентного прироста указанных показателей.

Рассмотрим в качестве примера динамику изменения общего числа выявленных заболевших (инфицированных) людей. Если разделить величину прироста этого показателя в какой-либо день на общее количество выявленных заболевших людей, зафиксированное в предыдущий день, и умножить полученное значение на 100, то получим процентный прирост общего числа инфицированных. Аналогично можно посчитать процентный прирост и для других групп. Обозначим значение процентного прироста выявленных заболевших в течение развития эпидемии от начала до дня T через $r(t)$, $t \in 1, 2, \dots, T$. Пусть временной горизонт развития эпидемии разбит на M промежутков $(T_{m-1}, T_m]$, $0 \leq T_{m-1} < T_m \leq T$, $m = 1, 2, \dots, M$. Тогда для любого промежутка $(T_{m-1}, T_m]$ и любого $k = 1, 2, \dots, (T_m - T_{m-1})$ динамику изменения количества инфицированных можно описать следующим образом:

$$I(T_{m-1} + k) = \left(1 + \frac{r(T_{m-1} + k)}{100} \right) I(T_{m-1} + k - 1), \quad k = 1, 2, \dots, (T_m - T_{m-1}).$$

Таблица. Прогнозы и фактические показатели

Дата	Описание прогноза	Фактические данные	Комментарий
5 мая	Россия может выйти на пик эпидемии (по количеству болеющих) ориентировочно 18 мая, когда скорость роста общего числа инфицированных будет около 2% и начнется снижение числа болеющих	20 мая процентный прирост опустился ниже нуля (-0.29%), однако затем начал возрастать и 26 мая достиг нового отрицательного значения -1.55%. Россия вышла на первый локальный пик эпидемии по количеству болеющих 25 мая - 230 996 человек	Ошибка прогноза дня выхода на пик эпидемии по количеству болеющих связана с нестабильной динамикой процентного прироста
16 мая	При динамике прироста, схожей с текущей, интервал снижения процентного прироста от 5 до 2-2.5% может составить не менее 12 дней. Россия может выйти на пик эпидемии по количеству болеющих ориентировочно 24 мая, достигнув уровня общего числа инфицированных 305 000-350 000 человек	Переход от уровня процентного прироста 5% до 2-2.5% занял 15 дней. Процентный прирост перешел рубеж 2.5% 27 мая, став равным 2.3%. Общее число инфицированных 24 мая оказалось равным 344 481 человек	Процентный прирост перешел рубеж 2.5% на один день раньше
31 мая	Общее число инфицированных к 10 июня может возрасти до 467 000 человек	Общее число инфицированных, зарегистрированное 10 июня, - 493 657 человек	Ошибка прогноза равна 5%
2 июня	роста выявленных заболевших от уровня 5% до уровня 2% занимает уже 21 день и может продлиться еще 1-2 дня, к 10-11 июня общее число инфицированных может увеличиться до 491 000 человек. К 20-21 июня их общее количество может возрасти до 569 000 человек	Общее число инфицированных, зарегистрированное 10 июня, - 493 657 человек. Фактическое количество инфицированных на 20 июня составило 576 952 человек	Ошибка прогноза на 10 июня составила 0.54%. Ошибка прогноза на 20 июня - 1.4%
15 июня	Общее количество заболевших к 20 июня будет около 582 000 человек, к 30 июня - около 655 000 человек	Фактически заболевших к 20 июня - 576 952 человек, к 30 июня - 647 849 человек	Ошибка прогноза на 20 июня - 0.9%. Ошибка прогноза на 30 июня - 0.17%

Для вычисления общего количества инфицированных в конце промежутка $(T_{m-1}, T_m]$ воспользуемся формулой

$$I(T_m) = \left(1 + \frac{r(T_{m-1} + 1)}{100}\right) \left(1 + \frac{r(T_{m-1} + 2)}{100}\right) \cdots \left(1 + \frac{r(T_m)}{100}\right) I(T_{m-1}).$$

В процессе построения прогнозируемой траектории изменения общего количества инфицированных на промежутке $(T_{m-1}, T_m]$ будем генерировать последовательность прогнозируемых значений процентного прироста $\tilde{r}(T_{m-1} + k)$, $k = 1, 2, \dots, (T_m - T_{m-1})$. Тогда прогнозируемую величину показателя выявленных заболевших $I(t)$ можно рассчитать так:

$$\tilde{I}(T_m) = \left(1 + \frac{\tilde{r}(T_{m-1} + 1)}{100}\right) \left(1 + \frac{\tilde{r}(T_{m-1} + 2)}{100}\right) \cdots \left(1 + \frac{\tilde{r}(T_m)}{100}\right) \tilde{I}(T_{m-1}).$$

При этом вместо величины $\tilde{I}(T_{m-1})$, характеризующей прогнозируемое значение показателя в конце предыдущего промежутка построения прогноза, может быть выбрана фактическая его величина, зафиксированная в конце данного промежутка, если это, по мнению экспертов, будет способствовать улучшению качества прогноза для следующего периода времени.

Учитывая выводы, приведенные в статье [6], аналогичную процедуру можно реализовать для группы прогнозирования значений $\tilde{A}(t)$ для группы A болеющих (активных случаев). Для построения прогноза динамики общего количества выздоровевших и умерших $\tilde{R}(t) + \tilde{F}(t)$ можно воспользоваться балансовым соотношением

$$I(t) = A(t) + R(t) + F(t), \quad t = 1, 2, \dots, T.$$

Построенная нами модель CBRR включает итеративную процедуру эвристического выбора длины промежутков, набора значений процентного прироста и других значимых параметров (пиков по приросту новых случаев и возможных периодов их достижения, пиков по количеству болеющих и др.). Эта процедура строится на основе информации о значениях аналогичных параметров при реализации предшествующих динамических процессов развития эпидемии. Важной ее составляющей является формирование цепочки ESC (Epidemic Spreading Chain) стран распространения эпидемии, включающей в себя несколько стран, упорядоченных по времени выхода их на одинаковые уровни значения выбранных параметров. Страна, для которой строится прогноз, называется страной-последователем (Country-Follower), остальные страны назовем странами-предшественниками (Country-Predecessor). При этом необходимо, чтобы в странах ESC были введены сравнительно одинаковые ограничения для распространения эпидемии (принятые решения о карантине, самоизоляции, социальной дистанции, блокировании транспортных потоков и т. п.). Например, эпидемия в России, стране-последователе, характеризуется более поздним по сравнению с другими странами моментом достижения процентным приростом одних и тех же значений. На основании этого в цепочку ESC при моделировании и прогнозировании динамики эпидемии в России в качестве предшественников включили Италию, Испанию, Великобританию и Францию. Последовательно сгенерированная на основании сделанного выбора траектория эволюции статистических данных об эпидемии, в частности, общего количества инфицированных людей, сравнивается с фактической статистикой.

На рис. 1 представлены графики спрогнозированной и фактической траекторий изменения общего числа инфицированных в период от 22 апреля до 30 июня 2020 г.

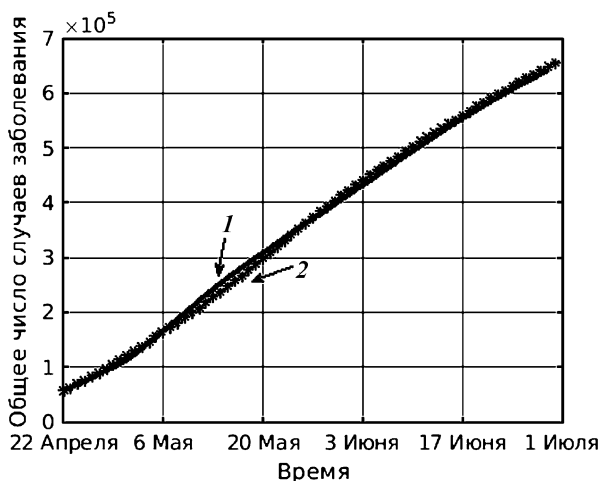


Рис. 1. Фактическая динамика и прогноз траектории изменения общего числа инфицированных на промежутке от 22 апреля до 30 июня 2020 г.
Общее число инфицированных: 1 – факт, 2 – прогноз.

Рисунок 2 демонстрирует динамику изменения процентного прироста общего числа случаев заболеваний и траекторию прогнозируемых значений этого показателя.

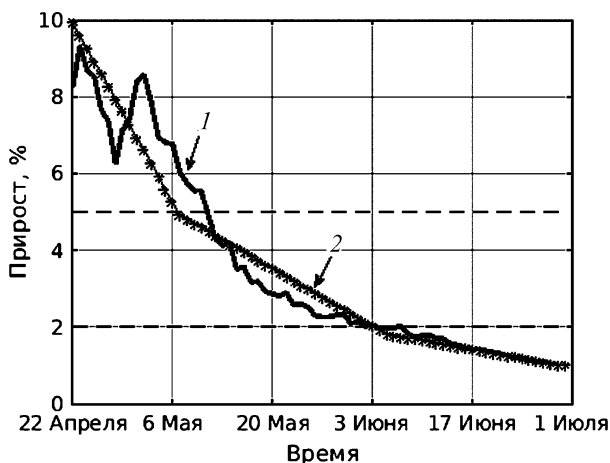


Рис. 2. Динамика изменения процентного прироста общего числа случаев заболеваний с 22 апреля по 1 июля 2020 г.
Изменение процентного прироста: 1 – факт, 2 – прогноз.

Траектории среднесрочного прогноза общего количества инфицированных и прогноза процентного прироста его и фактических показателей построены на промежутке от 23 апреля до 30 июня 2020 г. и представлены на рис. 1 и 2. Траектории среднесрочного прогноза объединяют три участка траекторий, последовательно построенных на промежутках краткосрочных прогнозов от 23 апреля до 7 мая, от 8 мая до 5 июня и от 6 июня до 30 июня, их длительность составляет 15, 28 и 25 дней

соответственно. Усредненное отклонение траектории прогноза общего количества инфицированных от фактической на первом промежутке равно -2.86% , на втором — 1.77% , на третьем — -0.13% , а для всего периода прогнозирования — 0.37% . В конце прогнозируемого периода 30 июня 2020 г. отклонение составляет -1.17% .

Прогнозирование пика эпидемии по количеству болеющих (активных случаев) имело свои особенности вследствие выхода России на три локальных пика: 25 мая — 230 996 болеющих, 1 июня — 234 146, 15 июня — 245 580. Задача прогнозирования пика эпидемии по количеству болеющих решалась в три этапа. Были последовательно построены краткосрочные прогнозы динамики количества активных случаев на четырех промежутках: от 23 до 30 апреля, от 1 до 19 мая, от 20 мая до 4 июня и от 5 до 22 июня. На первом промежутке усредненное отклонение прогнозируемой траектории от фактической составило 0.73% , на втором — 17.10% , на третьем — -0.9% и на четвертом — 6.02% . По прогнозу пика предполагалось, что он будет достигнут 4 июня 2020 г. при количестве активных случаев в этот день 238 260 человек. Фактически глобальный пик был достигнут 15 июня — 245 580 активных случаев. Ошибка прогноза составила 3% , а спрогнозированная дата выхода на пик оказалась внутри промежутка от 25 мая до 15 июня. Графики прогноза и фактической траектории активных случаев изображены на рис. 3.

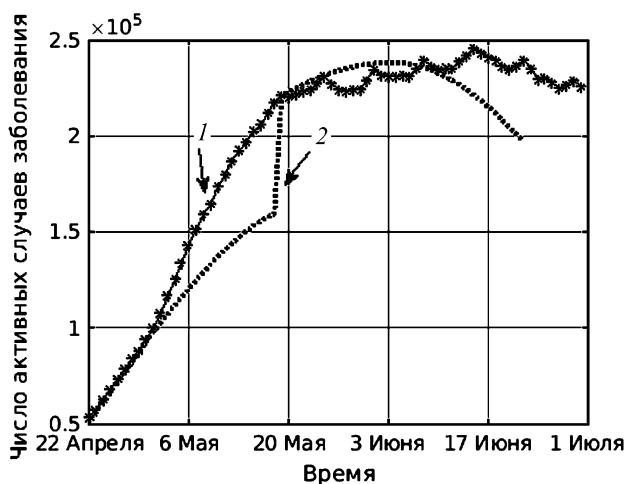


Рис. 3. Динамика активных случаев заражения: факт (1) и прогноз (2) в рассматриваемый период 2020 г.

4. Описание программного комплекса CBRR. Для программной реализации модуля CBRR была выбрана среда MATLAB, используемая исследователями в различных областях для моделирования динамических процессов. В качестве входных данных статистики зарегистрированных случаев заражения в разных странах использовали сведения из открытых источников, а именно, информацию, представленную на сайте Университета Джонса Хопкинса [12]. Эксперты Университета Джонса Хопкинса агрегируют данные из официальных источников различных стран для отслеживания распространения COVID-19. Данные обновляются в режиме реального времени каждый день. Процесс обработки и анализа данных для прогнозирования динамики эпидемии COVID-19 на основе метода прецедентов схематически изображен на рис. 4.

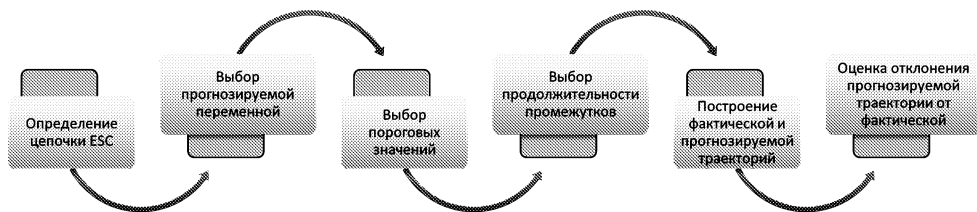


Рис. 4. Последовательность действий при построении прогноза динамики эпидемии на основе метода прецедентов

В начале пользователем формируется цепочка ESC из страны-последователя и стран-предшественников путем выбора стран из выпадающего списка. Далее производится выбор прогнозируемой переменной. В качестве альтернатив доступен выбор между числом выявленных случаев заболевания (I) и числом активных случаев (A). На следующем шаге выбираются пороговые значения процентного прироста. По умолчанию, пользователь может выбрать несколько значений от 20 до 0.2%. В конце подготовки данных для анализа выбирается продолжительность (в днях) промежутков перехода между установленными пороговыми значениями для прогнозируемой траектории. Соответствующий выбор доступен как в ручном, так и в автоматическом режимах (рис. 5). В последнем случае доступен выбор алгоритма построения траектории процентного прироста (пропорциональный, эффективный или среднепроцентный).

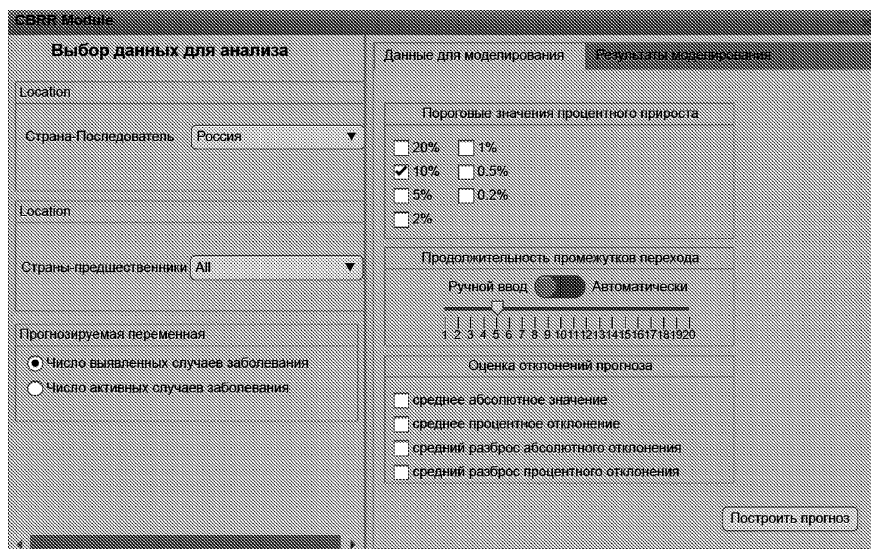


Рис. 5. Интерфейс диалогового окна модуля CBRR

На основании таких данных проводится построение фактической и прогнозируемой траекторий (для процентного и фактического прироста) на каждом промежутке прогнозирования и на временном интервале эволюции процесса изменения выбранной переменной от первой точки прогноза до конца текущего промежутка прогнозирования с выводом соответствующих графиков (рис. 6).

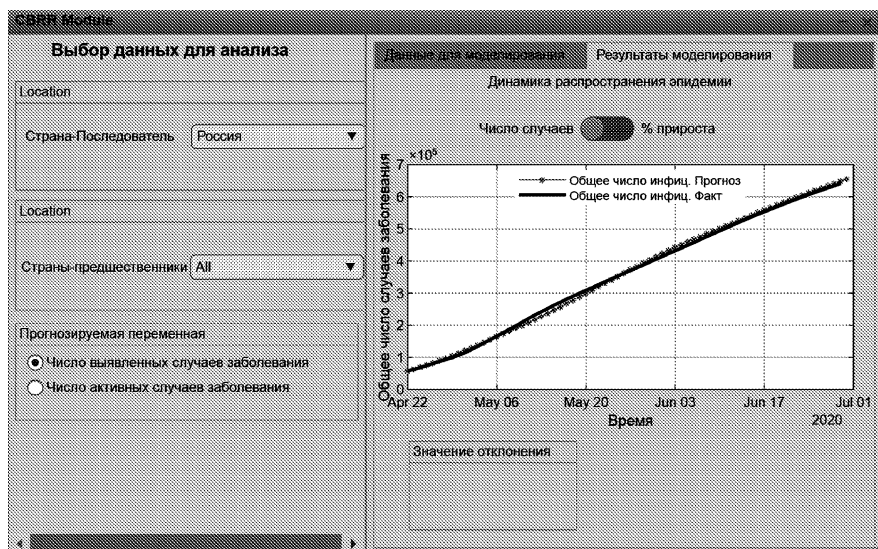


Рис. 6. Интерфейс окна с результатами моделирования в модуле CBRR

В качестве дополнительной информации для пользователя отображаются результаты расчета отклонения прогнозируемой траектории от фактической (среднее абсолютное значение, среднее процентное отклонение, средний разброс абсолютного отклонения, средний разброс процентного отклонения).

5. Заключение. Прогнозирование эпидемии нового коронавируса COVID-19 в России, впрочем, как и в других странах, — нестандартная задача. Системы дифференциальных и разностных уравнений разного типа, применяемые для прогнозирования развития эпидемии известного вируса, не вполне пригодны в этом случае, так как для их решения отсутствуют статистические данные из предыдущих периодов. Как показано в статье, метод прецедентов, использующий только информацию о развитии эпидемий, начавшихся в некоторых странах в более раннее время, может предложить вполне приемлемые прогнозы. К ключевым вопросам реализации метода относятся выбор цепочки стран-предшественников и использование технологий искусственного интеллекта, возможно машинного обучения, для построения зависимостей, необходимых для прогнозирования параметров, от имеющейся информации о развитии эпидемий в ближайшем прошлом. В связи с тем, что пандемия COVID-19 охватила более 200 стран мира, важной задачей является разработка программных комплексов для автоматизированного выбора цепочек стран-предшественников и реализации интеллектуальных алгоритмов прогнозирования развития эпидемий в многочисленных странах, оказывающихся в роли последователей.

Литература

1. Novel Coronavirus Global Research and Innovation Forum: Towards a Research Roadmap // WHO. URL: www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov (дата обращения: 15 июня 2020 г.).
2. Layne S. P., Nyman J. M., Morens D. M., Taubenberger J. K. New coronavirus outbreak: Framing questions for pandemic prevention // Sci. Transl. Med. 2020. Vol. 12. Iss. 534. N eabb1469. <https://doi.org/10.1126/scitranslmed.abb1469>

3. Wu J. T., Leung K., Leung G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study // *Lancet*. 2020. Vol. 395. Iss. 10225. P. 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
4. Models of Infectious Disease Agent Study // MIDAS Coordination Center. URL: www.midasnetwork.us/ (дата обращения: 15 июня 2020 г.).
5. Mandal M., Jana S., Nandi S. K., Khatua A., Adak S., Kar T. K. A model based study on the dynamics of COVID-19: Prediction and Control // *Chaos, Solitons and Fractals*. 2020. Vol. 136. N 109889. <https://doi.org/10.1016/j.chaos.2020.109889>
6. Fanelli D., Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France // *Chaos, Solitons and Fractals*. 2020. Vol. 134. N 109761. <https://doi.org/10.1016/j.chaos.2020.109761>
7. Bekirosab S., Kouloumpou D. SBDiEM: A new mathematical model of infectious disease dynamics // *Chaos, Solitons and Fractals*. 2020. Vol. 136. N 109828. <https://doi.org/10.1016/j.chaos.2020.109828>
8. Barnmparis G. D., Tsiornis G. P. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach // *Chaos, Solitons and Fractals*. 2020. Vol. 135. N 109842. <https://doi.org/10.1016/j.chaos.2020.109842>
9. Kondratyev M. A. Forecasting methods and models of disease spread // *Computer Research and Modeling*. 2013. Vol. 5. Iss. 5. P. 863–882.
10. Schmidt R., Waligora T. Influenza forecast: Case-Based Reasoning or statistics? // *Proceedings of the 11th International conference on knowledge-based intelligent information and engineering systems*. Pt I. Series Lecture Notes in Computer Science. 2007. Vol. 4692. P. 287–294.
11. Viboud C., Boelle P. Y., Carrat F., Valleton A. J., Flahault A. Prediction of the spread of influenza epidemics by the method of analogues // *American Journal of Epidemiology*. 2003. Vol. 158. Iss. 10. P. 996–1006.
12. Johns Hopkins Coronavirus Resource Center. URL: <https://coronavirus.jhu.edu/data> (дата обращения: 29 июня 2020 г.).

Статья поступила в редакцию 24 июля 2020 г.

Статья принята к печати 13 августа 2020 г.

Контактная информация:

Захаров Виктор Васильевич — д-р физ.-мат. наук, проф.; v.zaharov@spbu.ru

Балыкина Юлия Ефимовна — канд. физ.-мат. наук, доц.; j.balykina@spbu.ru

Predicting the dynamics of the coronavirus (COVID-19) epidemic based on the case-based reasoning approach

V. V. Zakharov, Yu. E. Balykina

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Zakharov V. V., Balykina Yu. E. Predicting the dynamics of the coronavirus (COVID-19) epidemic based on the case-based reasoning approach. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2020, vol. 16, iss. 3, pp. 249–259. <https://doi.org/10.21638/11701/spbu10.2020.303> (In Russian)

The case-based rate reasoning (CBRR) method is presented for predicting future values of the coronavirus epidemic's main parameters in Russia, which makes it possible to build short-term forecasts based on analogues of the percentage growth dynamics in other countries. A new heuristic method for estimating the duration of the transition process of the percentage increase between specified levels is described, taking into account information about the dynamics of epidemiological processes in countries of the spreading chain. The CBRR software module has been developed in the MATLAB environment, which implements the proposed approach and intelligent proprietary algorithms for constructing trajectories of predicted epidemic indicators.

Keywords: modeling, forecasting, COVID-19 epidemic, percentage rate of increase, case-based reasoning, heuristic.

References

1. Novel Coronavirus Global Research and Innovation Forum: Towards a Research Roadmap. *WHO*. Available at: www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov (accessed: June 15, 2020).
2. Layne S. P., Hyman J. M., Morens D. M., Taubenberger J. K. New coronavirus outbreak: Framing questions for pandemic prevention. *Sci. Transl. Med.*, 2020, vol. 12, iss. 534, no. eabb1469. <https://doi.org/10.1126/scitranslmed.abb1469>
3. Wu J. T., Leung K., Leung G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 2020, vol. 395, iss. 10225, pp. 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
4. Models of Infectious Disease Agent Study. *MIDAS Coordination Center*. Available at: <https://midasnetwork.us/> (accessed: June 15, 2020).
5. Mandal M., Jana S., Nandi S. K., Khatua A., Adak S., Kar T. K. A model based study on the dynamics of COVID-19: Prediction and Control. *Chaos, Solitons and Fractals*, 2020, vol. 136, no. 109889. <https://doi.org/10.1016/j.chaos.2020.109889>
6. Fanelli D., Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons and Fractals*, 2020, vol. 134, no. 109761. <https://doi.org/10.1016/j.chaos.2020.109761>
7. Bekirosab S., Kouloumpou D. SBDiEM: A new mathematical model of infectious disease dynamics. *Chaos, Solitons and Fractals*, 2020, vol. 136, no. 109828. <https://doi.org/10.1016/j.chaos.2020.109828>
8. Barmparis G. D., Tsironis G. P. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons and Fractals*, 2020, vol. 135, no. 109842. <https://doi.org/10.1016/j.chaos.2020.109842>
9. Kondratyev M. A. Forecasting methods and models of disease spread. *Computer Research and Modeling*, 2013, vol. 5, iss. 5, pp. 863–882.
10. Schmidt R., Waligora T. Influenza forecast: Case-Based Reasoning or statistics? *Proceedings of the 11th International conference on knowledge-based intelligent information and engineering systems. Pt I. Series Lecture Notes in Computer Science*, 2007, vol. 4692, pp. 287–294.
11. Viboud C., Boelle P. Y., Carrat F., Valleron A. J., Flahault A. Prediction of the spread of influenza epidemics by the method of analogues. *American Journal of Epidemiology*, 2003, vol. 158, iss. 10, pp. 996–1006.
12. *Johns Hopkins Coronavirus Resource Center*. Available at: <https://coronavirus.jhu.edu/data> (accessed: June 29, 2020).

Received: July 20, 2020.

Accepted: August 13, 2020.

Author's information:

Victor V. Zakharov — Dr. Sci. in Physics and Mathematics, Professor; v.zakharov@spbu.ru

Yulia E. Balykina — PhD in Physics and Mathematics, Associate Professor; j.balykina@spbu.ru