

ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ЭПИДЕМИЙ ВИРУСОВ НОВОГО ТИПА

Ю.Е. Балыкина

Санкт-Петербургский государственный университет
Россия, 199034, Санкт-Петербург, Университетская наб., 7-9
E-mail: j.balykina@spbu.ru

В.В. Захаров

Санкт-Петербургский государственный университет
Россия, 199034, Санкт-Петербург, Университетская наб., 7-9
E-mail: v.zaharov@spbu.ru

Ключевые слова: эпидемиология, модель CIRД, прогнозирование, модели эпидемий.

Аннотация: Повышение точности прогнозирования динамики эпидемий новых вирусов является актуальной общественно значимой проблемой, так как ее решение обеспечивает возможность для разработки региональными органами здравоохранения своевременных эффективных мероприятий по борьбе с распространением инфекций. В исследовании предлагается использование нового подхода к моделированию и прогнозированию динамики эпидемии вируса неизвестного типа на основе принципа динамического баланса и построении балансовой интегральной модели притока и оттока болеющих людей, отличной от модели SIR.

1. Введение

Исследование динамики развития инфекционных заболеваний является важной проблемой, так как ее решение обеспечивает возможность для разработки органами здравоохранения своевременных эффективных мероприятий по борьбе с распространением инфекций. При анализе и моделировании динамики распространения для новых вирусов основная проблема состоит в отсутствии наличия достаточного количества исторических данных для получения оценок значений входных параметров конкретных моделей. Учитывая максимальную неопределенность в параметрах распространения нового вируса, закономерно возникают трудности с оценкой потенциальной опасности и масштабов эпидемии. Традиционно используемые параметры в классической модели SIR (Susceptible – Infected – Recovered) не поддаются количественной оценке с достаточной точностью, что не позволяет устранить неопределенность будущей динамики распространения коронавируса [1]. Неполнота и неточность исходной информации для моделирования может заключаться: в невозможности полного сбора и учета информации о прогнозируемом процессе; в недостоверности и недостаточности исходной информации вследствие различной организации сбора данных, а также в возможном проявлении особенностей прогнозируемого процесса, существование которых не предполагалось. Более того, сама модель системы также меняется и может эволюционировать со временем (например, распространение новых штаммов вирусов). В таких случаях использование традиционных подходов к прогнозированию не обеспечивает получения адекватных результатов ввиду недостаточно обоснованного выбора меняющихся во времени параметров моделей. Таким образом, острой является проблема разработки новых

адаптивных методов прогнозирования динамических процессов с учетом неопределенностей в динамике развития системы.

Авторами был проведен систематический поиск исследований по моделям распространения динамики эпидемии COVID-19 с акцентом на первую волну, когда полная информация о новом коронавирусе отсутствовала. В целом существует четыре основных подхода к моделированию распространения инфекционных заболеваний [2]: модели на основе временных рядов, камерные модели, агентные (сетевые) модели и модели, построенные с использованием методов машинного обучения и эвристических подходов. Несколько исследователей использовали модель ARIMA для прогнозирования распространения новой пандемии коронавируса. Примерами таких исследований являются [3, 4]. В целом, хотя модели временных рядов являются популярным инструментом прогнозирования, использование этого подхода для оценки распространения новых инфекций имеет свои ограничения. В частности, отсутствие статистики за предыдущие периоды и, как следствие, неизвестные значения параметров не позволяют строить модели достаточной степени точности. Большая часть исследований берет за основу камерную модель SIR (Susceptible – Infected – Recovered) и ее вариации. Данная модель представляет собой систему дифференциальных уравнений, каждое из которых описывает динамику популяции в конкретной группе (камере). Базовая модель предполагает наличие трех групп - восприимчивых (S), инфицированных (I) и выбывших (R), при этом возможность повторного заражения в данной модели не учитывается. Впервые так называемая модель эпидемии SIR, описываемая системой трех обыкновенных дифференциальных уравнений, была представлена в работе Кермака и Маккендрика в 1927 году [5]. Вариации данной модели предполагают наличие большего числа камер, а также возможность возвращения в одно из состояний – например, модели SEIS и SEIR, в которых учитывается латентный период заболевания, когда человек не заразен, а также модель MSIR, в которой дети могут родиться с иммунитетом [6]. Большую популярность набирают модели, основанные на методах машинного обучения и дополнительных эвристиках. Например, в [7,8] анализируются возможности применения искусственных нейронных сетей для моделирования распространения COVID-19. В исследованиях [9, 10] рассматриваются возможности использования подхода на основе прецедентов (case-based reasoning), применения балансовой модели процентного прироста, а также интегральной модели притока и оттока [11].

2. Материалы и методы

Для прогнозирования динамики общего количества заболевших в регионе используется модель CBRR (Case-Based Rate Reasoning). Данная модель основана на подходе, включающем процедуры формирования для региона цепочки распространения вируса и основанном на методе прецедентов, который продемонстрировал высокую точность прогнозирования динамики этого показателя [9, 10]. В качестве фундаментального свойства эпидемиологического процесса используется свойство динамической сбалансированности значений общего количества заболевших в прошлые периоды и значений общего количества выздоровевших и умерших в текущий момент времени.

Важным этапом предложенного алгоритма в модели CBRR является формирование цепочки ESC (Epidemic Spreading Chain) регионов, упорядоченных по времени выхода процентного прироста зарегистрированных случаев заболевания на одинаковые уровни. Регион, для которого строится прогноз, называется регионом-последователем, остальные регионы называются регионами-предшественниками. В качестве регионов

могут выступать страны, или же отдельные их части (области, города и т.д.). Рассмотрим процесс построения цепочек распространения вируса в регионах на основе предложенного подхода.

Пусть $N = \{1, \dots, n\}$ – множество регионов. На основании имеющейся информации для каждого региона $i \in N$ можно вычислить значения процентного прироста общего количества заболевших $C_i(t)$ в этом регионе в день $t = 1, 2, \dots, T$ по формуле

$$r_{C_i}(t) = 100 \frac{C_i(t) - C_i(t-1)}{C_i(t-1)}.$$

Например, по имеющимся статистическим данным, в период первой волны эпидемии значения процентного прироста, начиная с уровня 10%, преимущественно имели тенденцию к убыванию. Конкретный состав цепочки распространения для каждого из регионов-последователей формируется экспертным путем. Формирование цепочек распространения для других значений пороговых уровней (5%, 2%, 1% и т.п.), происходит аналогичным образом. После построения набора допустимых цепочек распространения для региона-последователя при заданных пороговых уровнях процентного прироста появляется возможность оценить тренды снижения процентного прироста $r_C(t)$ на основе продолжительности снижения в регионах-предшественниках.

Оценка характеристики динамического баланса основывается на принципе динамического баланса эпидемиологического процесса, сформулированного в [10].

Принцип динамического баланса. Пусть заданы значения $t_0 \geq 0$ и $T > t_0$, такие, что $R(T) > C(t_0) > 1$. Тогда для решения $\tau(T)$ задачи (1)-(2)

$$(1) \quad \min_{t_0 \leq t \leq T} t,$$

$$(2) \quad C(t) \geq R(T),$$

значение $R(T)$ в день T удовлетворяет неравенствам:

$$(3) \quad C(\tau(T)) \geq R(T) \geq C(\tau(T) - 1).$$

Условие (3) определяет связь между числом выздоровевших и умерших людей в определенный день t и числом зафиксированных случаев $t - \tau(t)$ дней назад. Таким образом, с помощью условия (3) можно установить динамический баланс между значениями функций $R(t)$, $C(\tau(t) - 1)$ и $C(\tau(t))$, где $t - \tau(t)$ - характеристика динамического баланса эпидемиологического процесса. В условиях указанного принципа значение $R(T)$ будет принадлежать отрезку $[C(\tau(T) - 1), C(\tau(T))]$.

В этом случае общие балансовые уравнения для обобщенной модели CIR примут вид:

$$\begin{aligned} C(t) &= \left(1 + \frac{r_C(t)}{100}\right) \times C(t-1), \\ I(t) &= \left(1 + \frac{r_I(t)}{100}\right) \times C(t-1) - R(t), \\ R(t) &= \left(1 + \frac{r_R(t)}{100}\right) \times R(t-1), \\ R(t) &= \lambda_t C(\tau(t) - 1) + (1 - \lambda_t) C(\tau(t)). \end{aligned}$$

3. Результаты

Вычислительные эксперименты были проведены для России, а также для одного из крупнейших регионов – города Санкт-Петербург (рис. 1, 2).

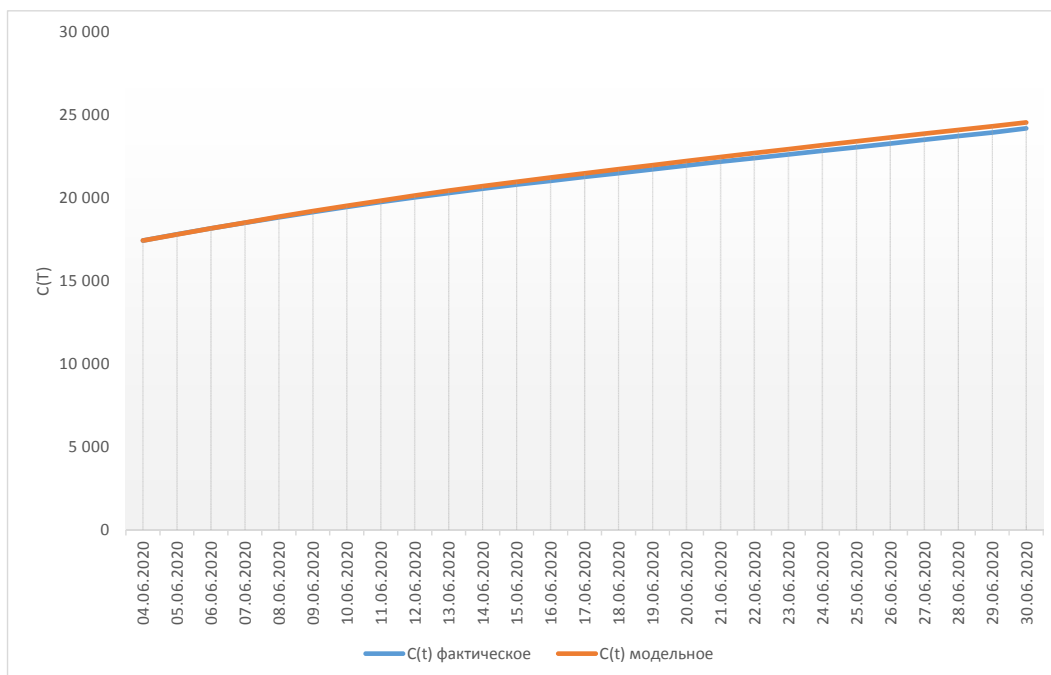


Рис. 1. Динамика общего числа случаев $C(t)$ в Санкт-Петербурге — фактическая и модельная траектории.

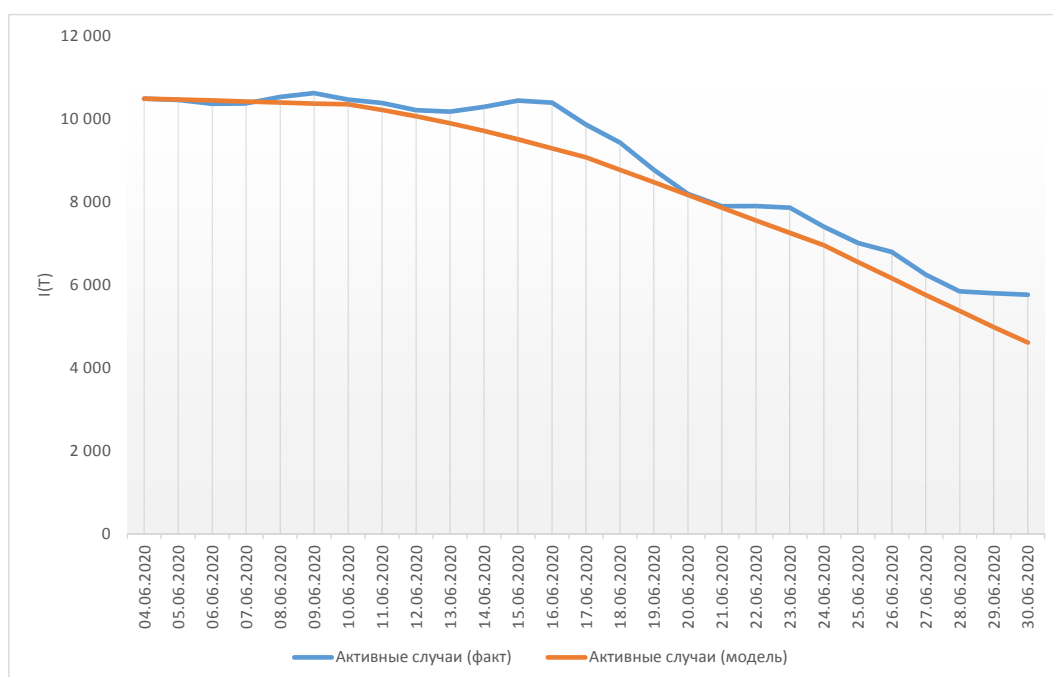


Рис. 2. Динамика общего числа случаев $I(t)$ в Санкт-Петербурге — фактическая и модельная траектории.

Отклонение смоделированной траектории от фактической траектории количества активных больных по показателям MAPE для России в целом составило 4,1%. При этом для динамики прогнозируемой траектории общего числа зарегистрированных случаев заболевания ошибка прогноза MAPE на рассматриваемом интервале составила 0,2%. Для динамики общего количества зарегистрированных случаев ошибка прогноза MAPE на рассматриваемом интервале с 6 по 30 июня 2020 г. составила 0,9% для Санкт-

Петербурга (рис. 1) и 0,2% для России. Отклонение расчётной траектории от фактической траектории числа активных больных для рассмотренного временного интервала 5.06.2020 – 30.06.2020 по показателю MAPE составило 5% для Санкт-Петербурга (рис.2) и 4,1% для России.

Исследование выполнено за счет гранта Российского научного фонда №23-21-10049, <https://rscf.ru/project/23-21-10049/>.

Список литературы

1. Zhang S., et al. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis // *Int. J. Infect. Dis.* 2020. Vol. 93. P. 201-204.
2. Shinde G.R., Kalamkar A.B., et al. Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art // *SN Computer Science*. 2020. Vol. 1, No. 197. doi: 10.1007/s42979-020-00209-9
3. Harvey A., Kattuman P. A farewell to R: time-series models for tracking and forecasting epidemics // *J. R. Soc. Interface*. 2021. Vol 18. P. 20210179. <http://doi.org/10.1098/rsif.2021.0179>
4. Aditya Satrio C.B., Darmawan W., et al. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET // *Procedia Comput. Sci.* 2021. Vol. 179. P. 524-532. doi: 10.1016/j.procs.2021.01.036
5. Anderson R.M., May R.M. *Infectious diseases of humans: Dynamics and control*. Oxford: Oxford University Press, 1991.
6. Csefalvay C. *Computational Modeling of Infectious Disease*, Academic Press, 2023. doi:10.1016/b978-0-32-395389-4.00012-8.
7. Malinzi J., Gwebu S., Motsa S. Determining COVID-19 Dynamics Using Physics Informed Neural Networks // *Axioms*. 2022. Vol. 11. No. 121. <https://doi.org/10.3390/axioms11030121>.
8. Xiao Ning, Linlin Jia, Yongyue Wei, Xi-An Li, Feng Chen, Epi-DNNs: Epidemiological priors informed deep neural networks for modeling COVID-19 dynamics // *Computers in Biology and Medicine*. 2023. Vol. 158. P. 106693. doi: 10.1016/j.combiomed.2023.106693.
9. Zakharov V., Balykina Y., Petrosian O., Gao H. CBRR Model for Predicting the Dynamics of the COVID-19 Epidemic in Real Time // *Mathematics*. 2020. Vol. 8. P. 1727. <https://doi.org/10.3390/math8101727>.
10. Захаров В.В., Балькина Ю.Е. Балансовая модель эпидемии COVID-19 на основе процентного прироста // *Информатика и автоматизация*. 2021. № 5 (20). С. 1034-1065. <https://doi.org/10.15622/20.5.2>.
11. Захаров В.В. Принцип динамического баланса демографического процесса и пределы роста населения Земли // *Докл. РАН. Матем., информ., проц. упр.* 2023. № 513. С. 108-114.