

Здравоохранение в информационном обществе**ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ ВИЗУАЛИЗАЦИИ ДАННЫХ
КОВИД-19 И ИЗУЧЕНИЯ КОНТРОЛЬНЫХ ПАРАМЕТРОВ МОДЕЛИ
РАСПРОСТРАНЕНИЯ ПАНДЕМИИ В РОССИИ И МИРЕ**

Статья рекомендована к публикации членом редакционного совета А.Н.Райковым 27.03.2022.

Левашкин Сергей Павлович

*Кандидат физико-математических наук, профессор
Поволжский государственный университет телекоммуникаций и информатики, НИЛ Искусственного интеллекта, заведующий лабораторией
Самара, Российская Федерация
serguei.levachkine@gmail.com*

Захарова Оксана Игоревна

*Кандидат технических наук, доцент
Поволжский государственный университет телекоммуникаций и информатики, НИЛ Искусственного интеллекта, заместитель заведующего лабораторией
Самара, Российская Федерация
zaharovaoksanaai@gmail.com*

Иванов Константин Николаевич

*Магистр
Поволжский государственный университет телекоммуникаций и информатики, НИЛ Искусственного интеллекта, инженер
Самара, Российская Федерация
tred1999@yandex.ru*

Аннотация

Представлена информационная система для визуализации данных о COVID-19 в регионах России и мира. Она включает: 1) адаптивно-компаративную многопараметрическую модель распространения эпидемии, являющуюся обобщением классических моделей SEIR; и 2) модуль визуализации и настройки параметров этой модели по эпидемиологическим данным, реализованный в приборной панели, называемой здесь «coVID». Данные для тестирования собираются в реальном времени ежедневно с марта 2020 года из открытых интернет-источников и размещаются на «ферме данных» (автоматизированная система сбора, хранения и предварительной обработки данных из разнородных источников), размещенной на удаленном сервере. Сочетание предложенной модели и дашборда дает возможность проводить наглядные численные эксперименты и сравнивать их с реальными данными, что позволяет максимально точно настроить параметры модели, превращая ее в интеллектуальную информационную систему поддержки принятия решений. Также определяются наиболее важные параметры модели.

Ключевые слова

информационные системы, большие данные, визуализация данных, оптимизация процесса принятия решений, дашборд, COVID-19, математические модели распространения эпидемии

© Левашкин С.П., Захарова О.И., Иванов К.Н., 2023

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2023_03_55

Введение

Сегодня человечество столкнулось с новым и ранее неизученным заболеванием – коронавирусной инфекцией (COVID-19). Болезнь распространилась так быстро и неожиданно, что эффективные средства реагирования, анализа и контроля не поспевали за скоростью распространения и эволюции эпидемии. Однако по мере прохождения первых волн заболевания постепенно проясняется как характер заболевания, так и необходимость соблюдения мер предосторожности, чтобы граждане могли обезопасить себя и свои семьи от угрозы распространения вируса. При этом от оперативных и скоординированных действий медицинских работников, мер безопасности, вводимых правительствами стран, реакции людей на эти меры, а также работы исследователей зависят человеческие жизни. На основе данных, опубликованных в ходе пандемии, важно проанализировать эволюцию заболевания, чтобы принять решения, которые повлияют на проблему увеличения случаев заболевания и смертности.

Чтобы проанализировать текущее и будущее состояние распространения пандемии, мы разработали информационную панель соVID для визуализации интернет-данных о COVID-19. Этот инструмент позволяет быстро отображать, а также управлять визуализацией собранной информации и анализировать развитие болезни, чтобы принимать решения, связанные с различными параметрами эпидемии (заболеваемость, смертность, вакцинация и т. д.). Сбор и анализ данных осуществляются по всем странам мира. Однако из-за ограниченного объема данной статьи мы в основном ориентируемся на данные по некоторым регионам РФ.

1 Адаптивно-компаративная модель распространения эпидемии и дашборд соVID

Панель инструментов основана на фреймворке Streamlit для объектно-ориентированного языка Python. Streamlit – это новейший подход к разработке пользовательских веб-приложений. Этот фреймворк сочетает в себе основы использования классических библиотек Python с их методами, функциями и аргументами, а также позволяет встраивать элементы HTML, CSS или JavaScript в исполняемый код. За визуализацию отвечают библиотеки Altair и Matplotlib. Для выполнения промежуточных преобразований используются следующие компоненты:

- NumPy — один из самых фундаментальных пакетов в Python. Он предоставляет высокопроизводительные объекты многомерного массива и инструменты для работы с ними. Кратко его можно охарактеризовать как эффективный контейнер универсальных многомерных данных.
- Pandas — пакет с открытым исходным кодом, который предоставляет высокоэффективные, простые в использовании структуры данных и инструменты.
- PIL — модуль для работы и сохранения изображений в различных форматах.
- Scipy — используется для выполнения научных и инженерных вычислений.
- Math — этот модуль предоставляет обширный функционал для работы с числовыми данными.

1.1 Структура дашборда

Структура дашборда представляет собой одностраничный сайт, разделенный на логически определенные блоки, содержащие следующую информацию:

1. Данные по странам о заболеваниях, выздоровлениях, смертях, прививках и т.д.
2. Региональная ситуация по эпидемиологическим показателям в субъектах Российской Федерации и динамика волн распространения болезни.
3. Внизу страницы находится информация о правообладателе панели инструментов.

Визуализированная информация отображается в виде графиков временных рядов и диаграмм, удобных для анализа. Сравнить ситуацию в разных странах и регионах РФ можно как по результатам работы нашей модели, так и по реальным данным. Встроенная функциональность упомянутых выше библиотек Python позволяет удобно экспортировать графики в различные форматы.

1.2 Работа с приборной панелью

При разработке дашборда мы изучили данные о COVID-19, предоставленные рядом крупных организаций по всему миру. Этими организациями являются Университет Джона Хопкинса (JHU) (<https://coronavirus.jhu.edu/map.html>), Всемирная организация здравоохранения (WHO) (<https://covid19.who.int/>) и многие другие [5]. Они сообщают актуальные данные, но не дают

возможности напрямую проводить исследования на своих платформах и управлять параметрами той или иной модели пандемии. При этом нашей основной целью является максимально точная настройка контрольных параметров модели распространения пандемии, проведение численных экспериментов по выявлению наиболее значимых параметров и разработка инструмента оптимизации эпидемического прогноза с целью минимизации ущерба от болезни.

На рисунке 1 показаны основные процессы, реализованные в информационной системе, связанные с разработанной средой сбора и обработки данных. Сбор и обработка данных происходят на удаленном сервере; агрегированный вывод пакетов данных экспортируется в каталог GitHub для дальнейшей визуализации. В качестве хостинга веб-приложения был выбран ресурс Heroku. Доступ к панели можно получить по следующему адресу: <https://dashai.herokuapp.com/>

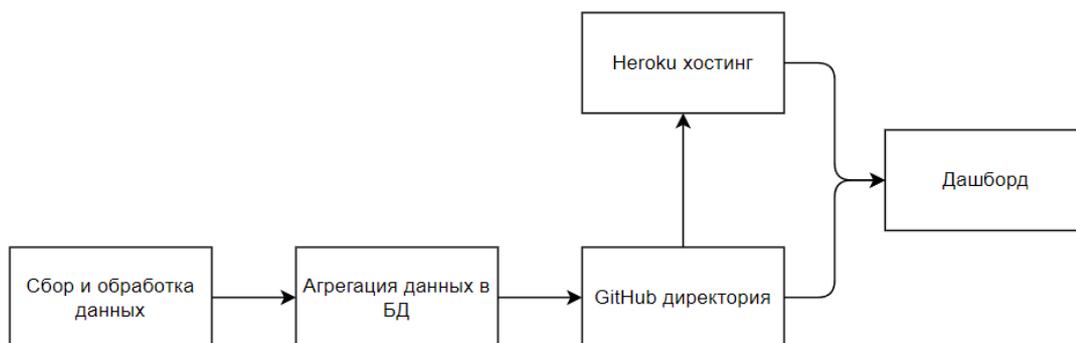


Рисунок 1. Структура дашборда

Рассмотрим возможности дашборда более подробно. В первом блоке представлена информация о количестве случаев заболевания, выздоровлений и смертей за время эпидемии в разных странах мира с возможностью сравнения этих показателей: линейный график (рисунок 2) и гистограмма (рисунки 3 и 4).

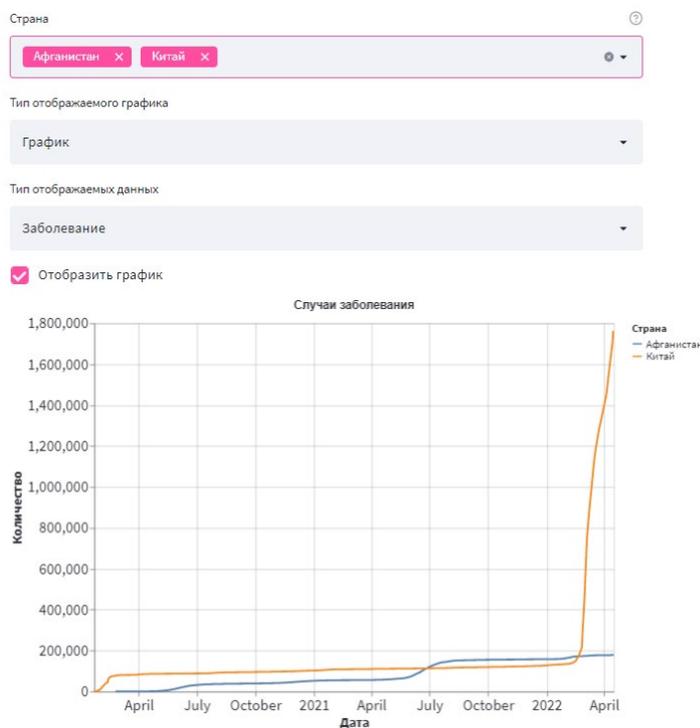


Рисунок 2. Случаи заболевания по странам (линейный график)

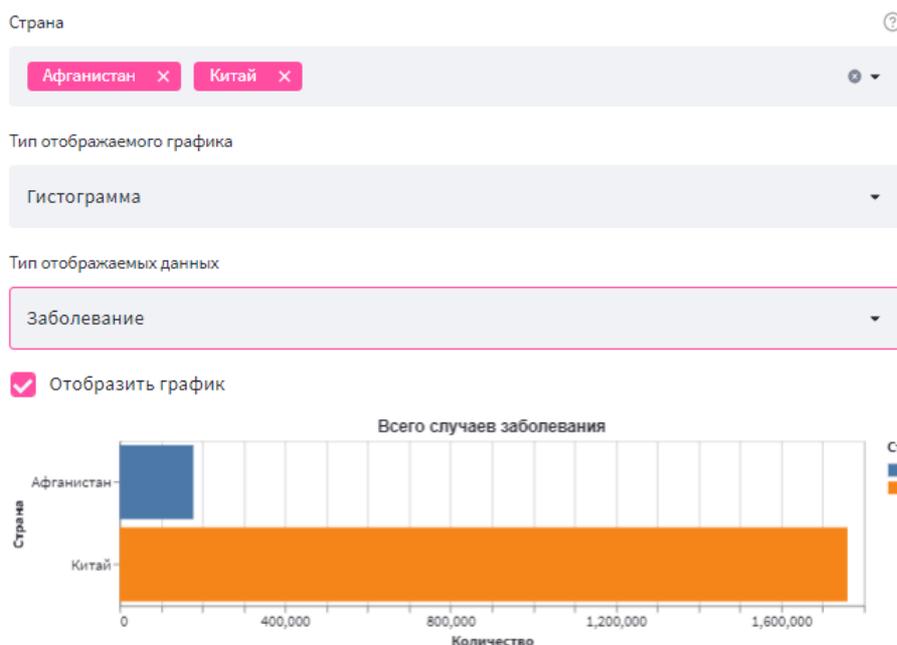


Рисунок 3. Случаи заболевания по странам (гистограмма)

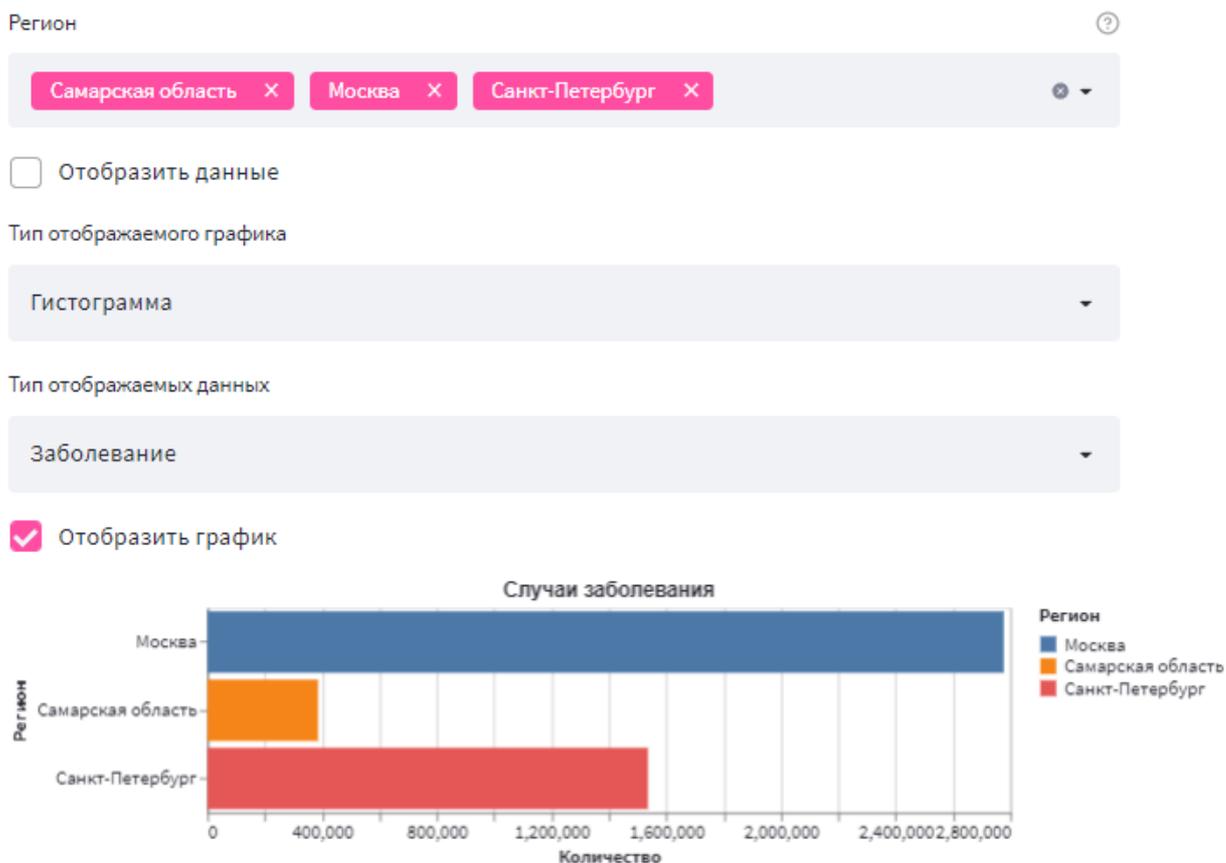


Рисунок 4. Случаи заболевания по регионам России (гистограмма)

На панели также отображаются фрагменты наборов данных (рисунок 5) и их визуализация (рисунок 6).

Регион



Самарская область X



Отобразить данные

	Дата	Заболевание	Смерти	Выздоровление	Регион	Co
38	2020-03-27	12.0000	0.0000	0.0000	Самарская область	RU
96	2020-03-28	14.0000	0.0000	1.0000	Самарская область	RU
157	2020-03-29	14.0000	0.0000	1.0000	Самарская область	RU
222	2020-03-30	14.0000	0.0000	1.0000	Самарская область	RU
290	2020-03-31	15.0000	0.0000	1.0000	Самарская область	RU
363	2020-04-01	15.0000	0.0000	2.0000	Самарская область	RU
437	2020-04-02	17.0000	0.0000	2.0000	Самарская область	RU
512	2020-04-03	17.0000	0.0000	7.0000	Самарская область	RU
589	2020-04-04	17.0000	0.0000	7.0000	Самарская область	RU
667	2020-04-05	17.0000	0.0000	8.0000	Самарская область	RU

Рисунок 5. Отображение фрагмента набора данных по Самарской области России

Тип отображаемого графика

График

Тип отображаемых данных

Выздоровление

Отобразить график

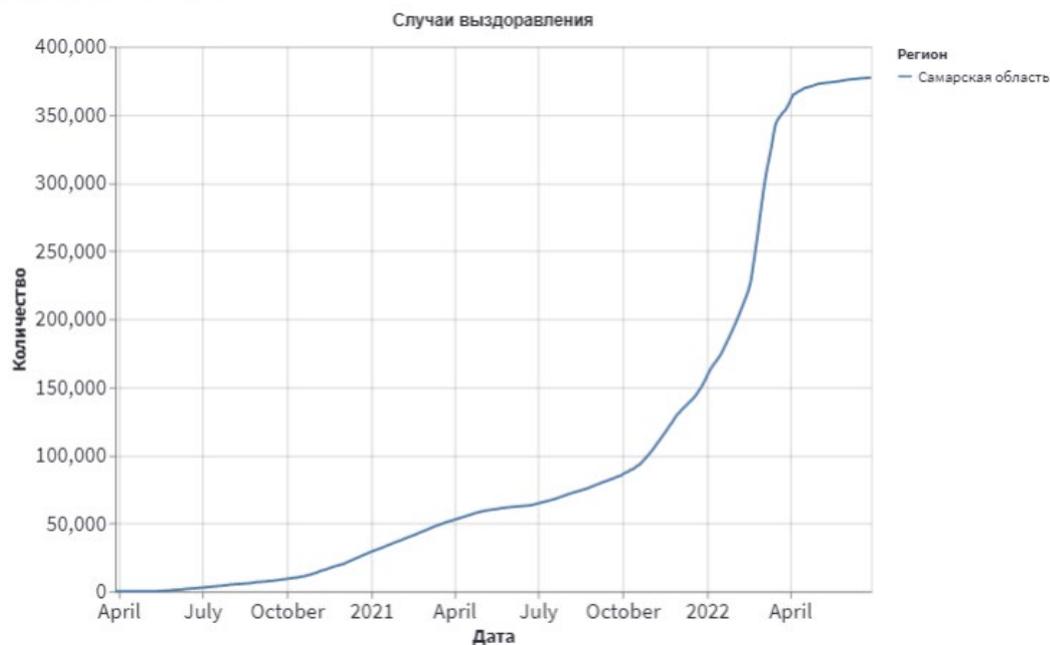


Рисунок 6. Визуализация собранных и обработанных данных в регионах РФ

Динамика распространения эпидемии носит волнообразный характер. Ее показатели связаны с мерами и ограничениями, вводимыми государством, а также с реакцией населения на введенные ограничения, которые являются основными управляющими параметрами модели.

Вариант сглаживания связан с так называемым «инкубационным периодом». Инкубационный период – это период между заражением и появлением клинических симптомов заболевания. По данным ВОЗ, в случае коронавирусной инфекции продолжительность инкубационного периода составляет 1-14 дней (см., например, <https://pubmed.ncbi.nlm.nih.gov/32150748/>). Используя эту концепцию, мы генерируем дифференцированные гладкие кривые.

Реализация этой процедуры выглядит следующим образом:

```
# Smooth option
train_split = 1
timelen = int(data.shape[0] * train_split)
t = np.arange(timelen)
ill_time = st.slider("Smoothing option", min_value=7, max_value=21,
value=14, step=7, key=15)
```

Дашборд позволяет отслеживать эту динамику в регионах РФ (рисунок 7) и в других странах (рисунки 8 и 9) по ряду показателей.

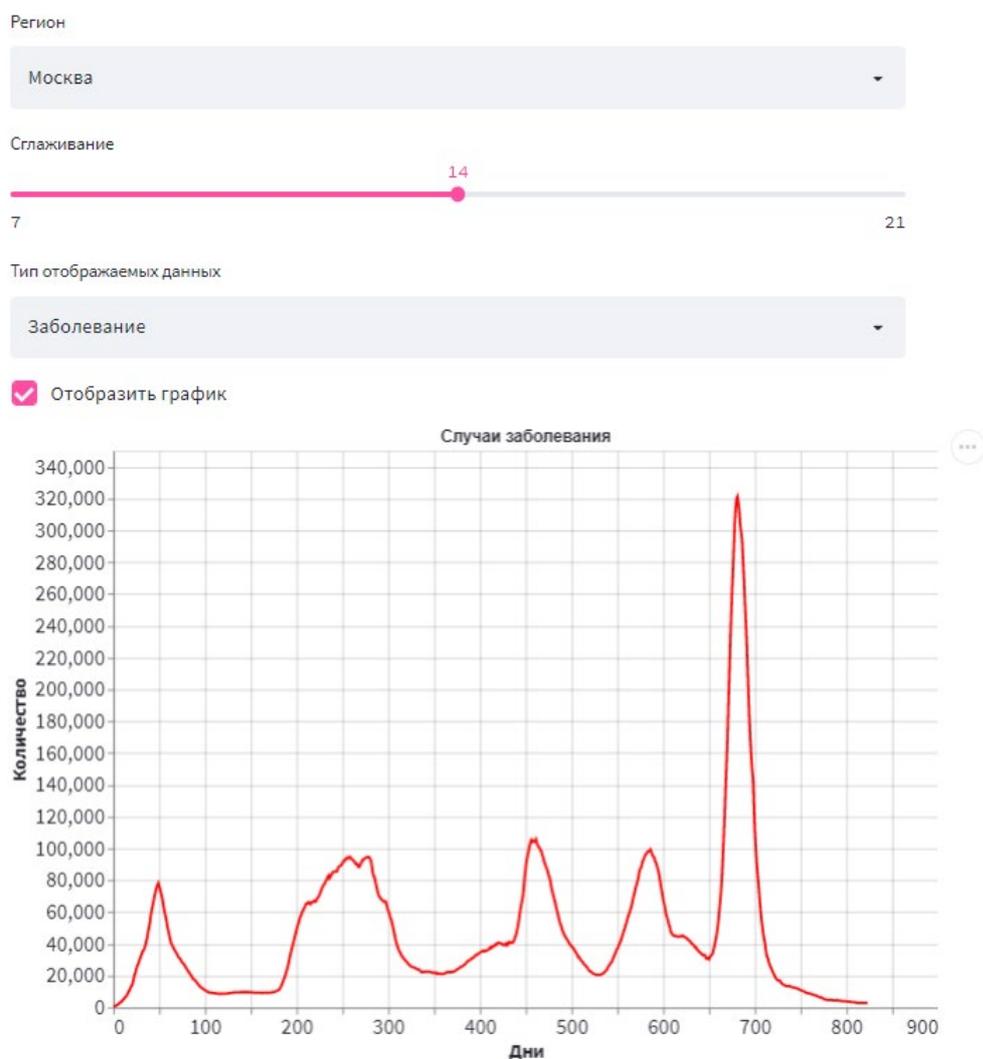


Рисунок 7. Динамика распространения заболевания в Москве с начала эпидемии

Визуализация динамических данных по странам

Выберите настройки для отображения данных

Страна

Россия

Сглаживание

7 14 21

Тип отображаемых данных

Заболевание

Отобразить график

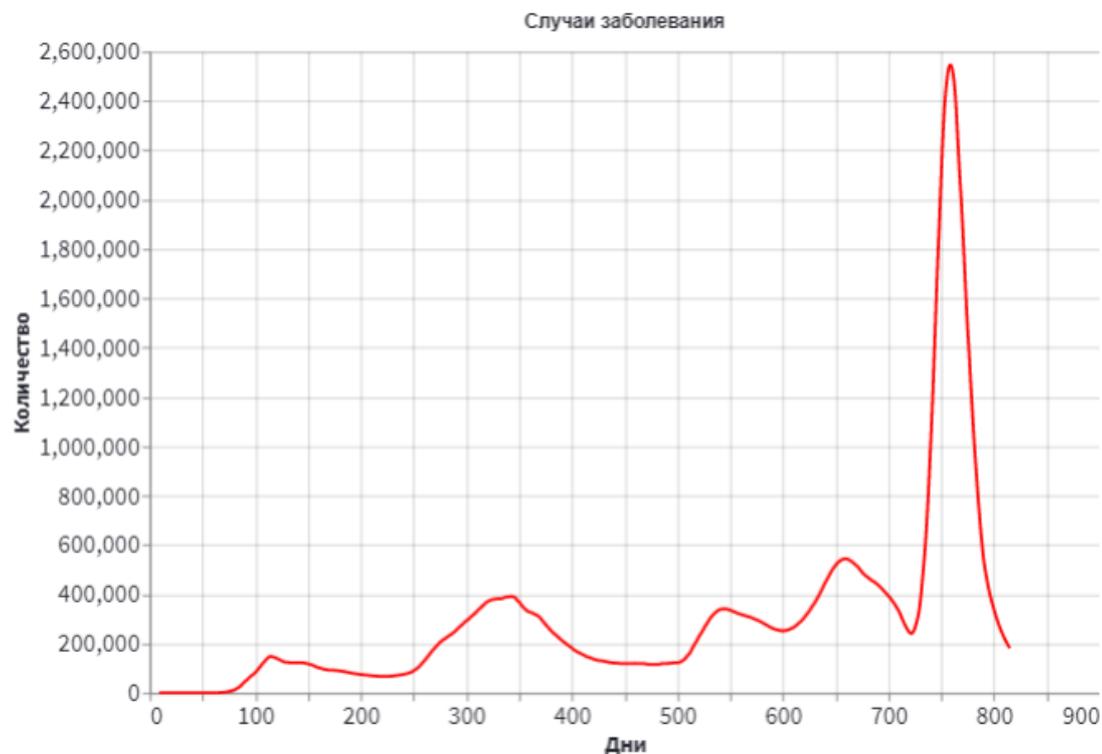


Рисунок 8. Динамика распространения заболевания в России с начала эпидемии

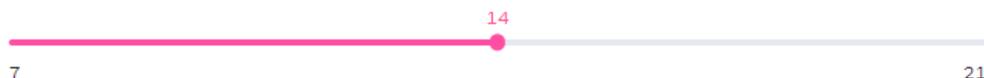
Визуализация динамических данных по странам

Выберете настройки для отображения данных

Страна

Беларусь

Сглаживание



Тип отображаемых данных

Заболевание

Отобразить график

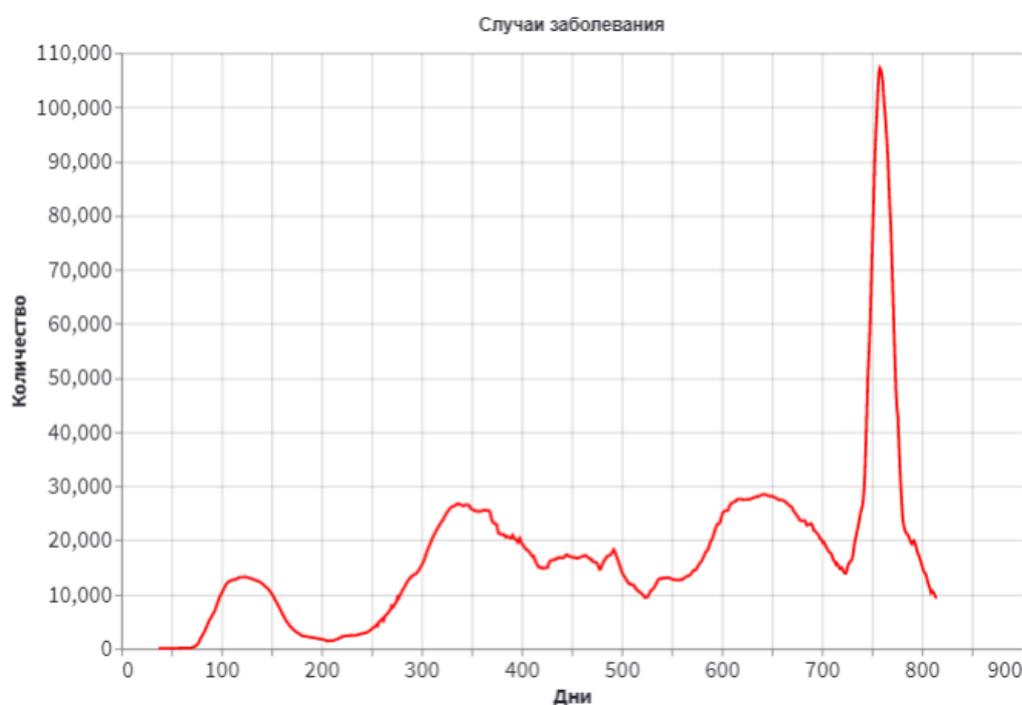


Рисунок 9. Динамика распространения заболевания в Беларуси с начала эпидемии

2 Адаптивно-компаратментная модель распространения эпидемии

На сегодняшний день известны и общепризнаны множество математических моделей распространения эпидемий (см., например, [1-6]). Большинство из этих моделей рассматривают основные стадии/состояния инфекционного процесса типа SEIR. SEIR представляет собой основные показатели течения пандемии в человеческом сообществе: S - восприимчивый, E - в инкубационном периоде (латентный), I - инфицированный, R - выздоровевший. Подробно рассмотрев эти модели, мы заметили, что для каждой из них характерно последовательное добавление новых стадий/состояний инфекционного процесса для более точного описания реальной динамики распространения эпидемии. Модели составляют эпидемиологические прогнозы, выполняя их на различные периоды в зависимости от целей исследования. Таким образом, построив краткосрочный прогноз на несколько недель вперед, можно составить план

применения оперативного управления для выявления эпидемических вспышек заболеваемости не только в отдельных регионах, но и в целых странах.

Пандемию COVID-19 можно однозначно рассматривать как серьезную биогенную угрозу, наносящую ощутимый ущерб социально-экономическим системам, государственным образованиям и обществу. При этом сверхактуальными научными задачами являются корректировка существующих и разработка новых математических моделей распространения эпидемий и методов информационного мониторинга не только распространения вируса, но и социально-экономической среды. И те, и другие должны отражать современные реалии и максимально точно описывать как саму биогенную угрозу, так и ее последствия. Поэтому своевременный мониторинг экономических, психологических и других социальных процессов, происходящих в «вирусной среде», позволит принимать более оптимальные управленческие решения и минимизировать последствия пандемии за счет более точной настройки модели прогноза.

Наш подход основан на модели SEIR, которая претерпела значительные изменения за счет добавления ряда новых параметров и процессов. Кроме того, параметры модели определяются и корректируются по данным о распространении пандемии в том или ином регионе, собираемым и структурируемым автоматически из множества различных интернет-источников. Для этого была разработана и внедрена так называемая «ферма данных» [5]. В результате получается адаптивно-компаратментная модель, которая характеризуется возможностью исследования проблемы с использованием современных подходов к оптимизации и интеллектуальному анализу данных. При разработке модели мы исходили из соображений оптимального ограничения количества управляющих параметров для наиболее точного описания имеющихся данных и построения краткосрочных и среднесрочных прогнозов.

Научная новизна данного исследования заключается в следующем:

- 1) Интеграция системы мониторинга и модели распространения вирусных эпидемий;
- 2) Использование информационной панели для настройки параметров модели на данных о COVID-19, полученных из ограниченного набора предопределенных интернет-источников;
- 3) Выявление слабо предсказуемой реакции общества на те или иные события, выраженные интернет-контентом.

Адаптивно-компаратментно-прогнозная модель на реальных данных о COVID-19 с описанием параметров управления модели выглядит следующим образом [5]:

$$\frac{dS}{dt} = -\beta(t) \frac{SI}{N}, \quad (1)$$

$$\frac{dE}{dt} = \beta(t) \frac{SI}{N} - \sigma E, \quad (2)$$

$$\frac{dI}{dt} = \sigma E - (1 - \alpha)\gamma I - \alpha R_0 I \quad (3)$$

$$\frac{dR}{dt} = (1 - \alpha)\gamma I \quad (4)$$

$$\frac{dD}{dt} = \alpha R_0 I \quad (5)$$

где:

$\beta(t) = \beta_0 * g(t) * s(t)$ – функция изменения базового коэффициента скорости передачи инфекции β_0 (или просто – скорость передачи инфекции);

$s(t) = \frac{1}{1 + e^{c_1(s_i(t) - c_2)}}$ – функция преобразования индекса самоизоляции от Яндексa (yandex.ru),

$s_i(t)$ берется из данных; значения c_1 и c_2 приведены в таблице 1;

$g(t) = \begin{cases} g_1, & \text{if } t < t_g \\ g_2, & \text{if } t > t_g \end{cases}$ – функция управления со стороны государства (карантин и другие ограничительные меры);

t – время;

- N – общая численность населения;
- σ – скорость передачи инфекции из группы (E) в группу (I);
- γ – скорость восстановления;
- α – уровень смертности (% от числа инфицированных);
- $R_0 - 1$ / время нахождения в критическом состоянии перед смертью.

Для определения интервалов оптимизируемых параметров мы проанализировали ряд текущих (2022 г.) исследований моделей COVID-19 и оценок, полученных из последних отчетов ВОЗ. Результаты этого анализа показаны в следующей таблице 1. Результаты, полученные с помощью модели (1)–(5), встроены в информационную панель coVID.

Таблица 1. Диапазоны эпидемических параметров

Параметр	Диапазон
β_0	0.2–3.5
c_1	0–1.0
c_2	0–5.0
σ	0–1.0
γ	0.1–0.2
α	0–1.0
R_0	0–1.0
tg	40–130.0
g_2	0–1.0
g_1	0–1.0

3 Результаты

На дашборде представлены результаты работы ранее описанной адаптивно-компарментной модели распространения эпидемии коронавируса в регионах Российской Федерации. Ключевой особенностью является возможность взаимодействия с управляющими параметрами модели, влияющими на изменение показателей модели. Пользователи дашборда могут настраивать ключевые параметры модели, чтобы отслеживать изменения в поведении результатов модели по отношению к реальным данным и, таким образом, получать более точные прогнозы. Этот процесс осуществляется путем вызова специального меню (рисунки 10, 11). Дашборд имеет удобный экспорт в формате JSON как исходных параметров, так и параметров, которые изменились (рисунок 12).

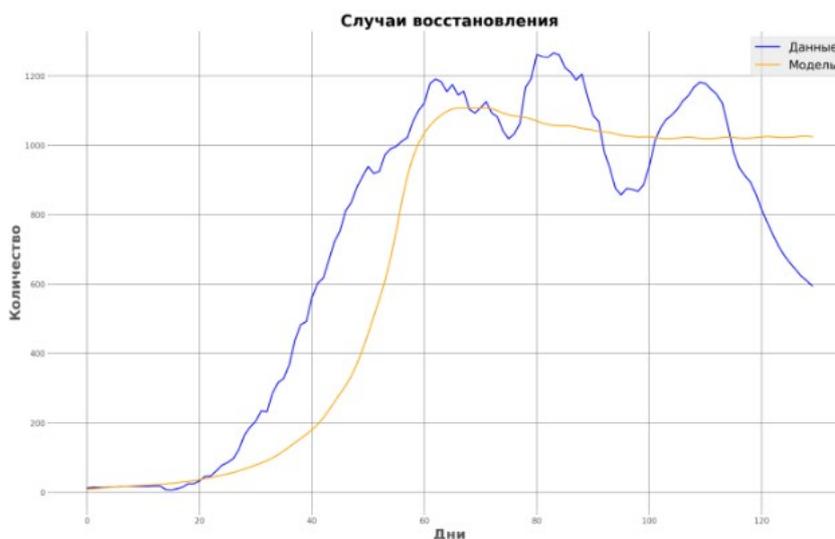


Рисунок 10. Фрагментарный вывод результатов адаптивно-компарментной модели

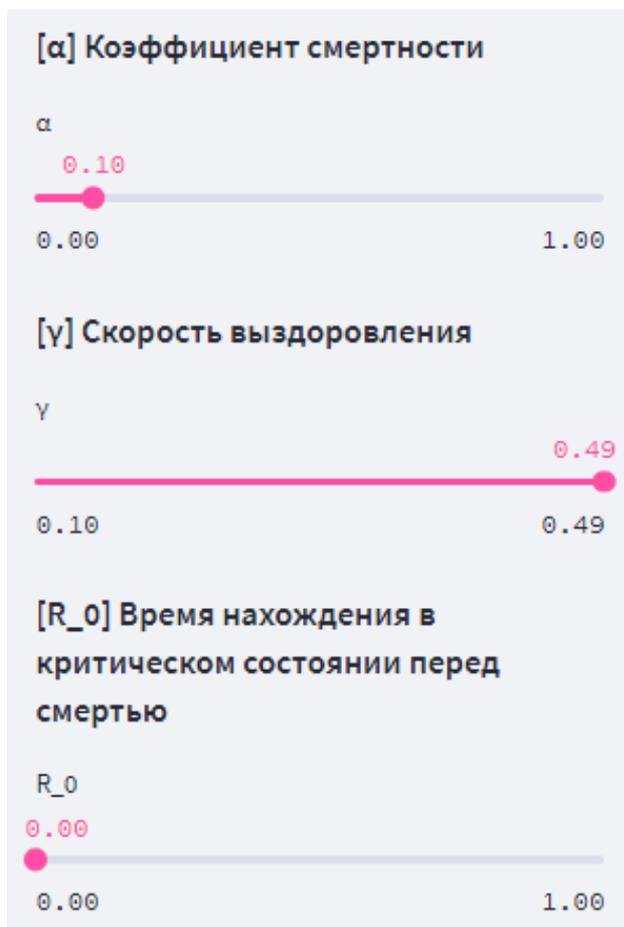


Рисунок 11. Фрагмент меню настройки параметров управления

```

{
  "beta_0" : 0.486821
  "c_1" : 0.88213
  "c_2" : 2.18778
  "sigma" : 0.198157
  "alpha" : 0.102617
  "gamma" : 0.490706
  "R_0" : 0.00134632
  "tg" : 55.6341
  "g_1" : 0.844033
  "g_2" : 0.117687
}

```

Рисунок 12. Вывод параметров модели в формате JSON

Используя дашборд, мы можем сравнить графики, построенные на реальных данных, собранных из разных интернет-ресурсов, и графики, полученные на выходе модели (1-5) (рисунок 13).

Тип отображаемых данных

Выздоровление

Отобразить график

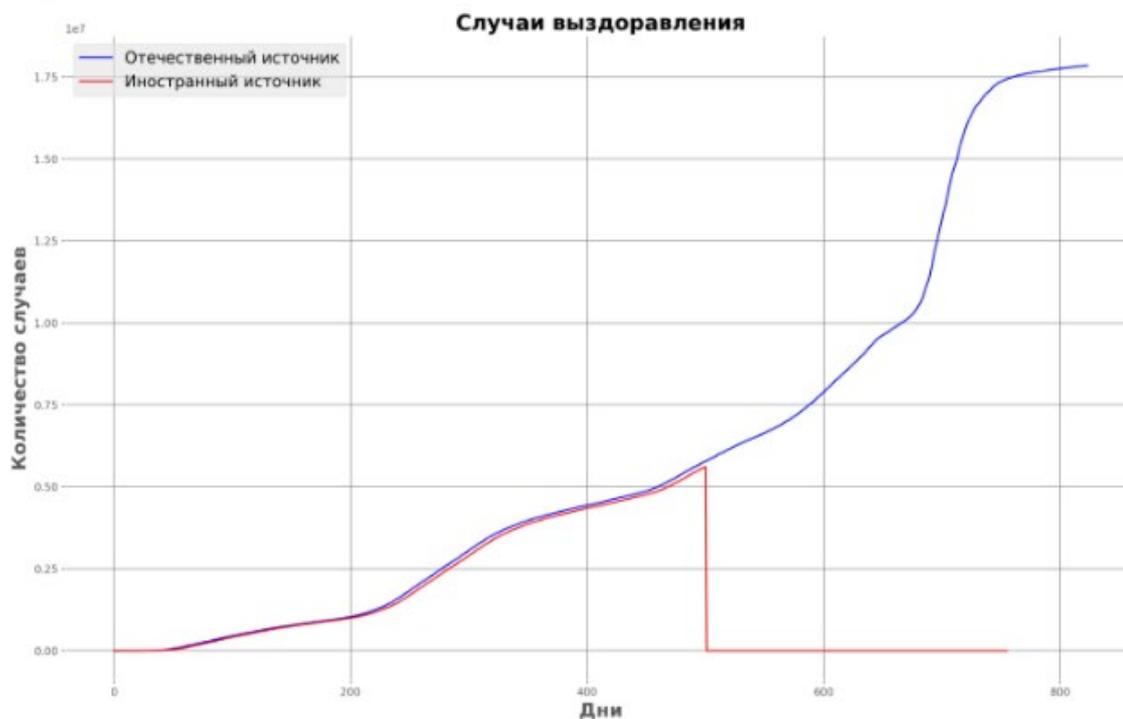


Рисунок 13. Сравнение данных разных источников о случаях выздоровления в России. Интересно отметить, что JHU перестал публиковать случаи выздоровления примерно на 500-й день эпидемии, но после этого публиковал актуальные данные о случаях заражения и смерти (см. эти графики на панели инструментов <https://dashai.herokuapp.com/>)

4 Заключение

Разработанная панель мониторинга распространения COVID-19 позволяет понять поведение и эволюцию этого заболевания за счет визуализации реальных данных (с помощью «фермы данных») и сравнения их с результатами предложенной адаптивно-компаративной модели. Кроме того, такой подход позволяет не только оценить фактическое количество заболеваний, смертей и другие параметры, но и определить страны и регионы РФ, наиболее пострадавшие от пандемии, и сравнить медицинские показатели эпидемии. Этот инструмент также позволяет изучить параметры модели и их влияние на постоянно меняющуюся ситуацию с распространением пандемии. Используя нашу информационную панель соVID, различные организации могут принимать превентивные меры для минимизации воздействия эпидемии. Находясь в активной разработке данного дашборда, мы уверены, что его использование в практических целях позволит выявить новые аспекты течения этого заболевания и разработать эффективный инструмент управления параметрами заболевания. Также будучи наглядным и простым в использовании, он найдет очень широкую аудиторию пользователей, в том числе среди неспециалистов в области информационных технологий, к которым в первую очередь относятся руководители различных государственных и негосударственных организаций.

Благодарности

Работа выполнена при частичной поддержке РФФИ, грант № 20-04-60160_Вирусy.

Литература

1. Н. Бейли. Математика в биологии и медицине. М.: МИР, 1970. 326 с.
2. Б. Боев. Современные этапы математического моделирования процессов развития и распространения инфекционных заболеваний. В.: Эпидемиологическая кибернетика: модели, информация, эксперименты. 1991. С. 6–13.
3. М. Кондратьев. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование, 2013. Т. 5. № 5. С. 863–882.
4. Дж. Вандер Плас. Python. Анализ данных и машинное обучение. СПб.: Питер, 2018. 901 с.
5. S. Levashkin, O. Zakharova, S. Agarov, K. Ivanov, E. Kuzmina, V. Sokolovsky, A. Monasova, A. Vorobyov, D. Apeshin. Study of SEIRD adaptive-compartmental model of COVID-19 epidemic spread in Russian Federation using optimization methods // Mathematical biology and bioinformatics, 2021. V. 16. No. 1. Pp. 136-151.
6. O. Krivorotko, S. Kabanikhin – Mathematical models of COVID-19 spread. 2022. URL: <https://arxiv.org/abs/2112.05315>

INFORMATION SYSTEM FOR COVID-19 DATA VISUALIZATION AND STUDY OF CONTROL PARAMETERS OF THE PANDEMIC SPREAD MODEL IN RUSSIA AND ABROAD

Levashkin, Sergei Pavlovich

*Candidate of Physical and Mathematical Sciences, Ph.D, Professor
Federal State Budgetary Educational Institution of Higher Education "Volga State University of Telecommunications and Informatics", Scientific Research Laboratory of Artificial Intelligence, Head of Laboratory
Samara, Russian Federation
serguei.levachkine@gmail.com*

Zaharova, Oksana Igorevna

*Candidate of Technical Sciences, Associate Professor
Federal State Budgetary Educational Institution of Higher Education "Volga State University of Telecommunications and Informatics", Scientific Research Laboratory of Artificial Intelligence, deputy head of the laboratory
Samara, Russian Federation
zaharovaoksanaai@gmail.com*

Ivanov, Konstantin Nikolaevich

*Master
Federal State Budgetary Educational Institution of Higher Education "Volga State University of Telecommunications and Informatics", Scientific Research Laboratory of Artificial Intelligence, engineer
Samara, Russian Federation
tred1999@yandex.ru*

Abstract

We develop an information system for COVID-19 data visualization in the regions of Russia and the world. It includes: 1) an adaptive-compartmental multi-parametric model of the epidemic spread, which is a generalization of the classical SEIR models; and 2) a module for visualizing and setting the parameters of this model according to epidemiological data, implemented in a dashboard, called herein "coVID". Data for testing have been collected since March 2020 on a daily basis from open Internet sources and placed on a "data farm" (automated system for collecting, storing and pre-processing data from heterogeneous sources) hosted on a remote server. The combination of the proposed model and the dashboard gives the ability to conduct visual numerical experiments and compare them with real data allowing most accurately tune the model parameters thus turning it into an intelligent information system to support a decision-making. The most important model parameters are also determined.

Keywords

information systems, big data, data visualization, optimization of the decision-making process, dashboard, COVID-19, epidemic spread mathematical models

References

1. N. Beili. Matematika v biologii i medicine. M.: MIR, 1970. 326 p.
2. B.Boev. Sovremennye etapy matematicheskogo modelirovaniya processov razvitiya i rasprostraneniya infekcionnyh zabolevanii. V.: Epidemiologicheskaya kibernetika: modeli, informacia, eksperimenty. 1991. P. 6-13.
3. M. Kondrat'ev. Metody prognozirovaniya i modeli rasprostraneniya zabolevanii // Comp'uternye issledovaniya i modelirovanie, 2013. T. 5. № 5. P. 863-882.
4. J. Vander Plas. Python. Analiz dannyh i mashinnoe obuchenie. SPb.: Piter, 2018. 901 p.
5. S. Levashkin, O. Zakharova, S. Agapov, K. Ivanov, E. Kuzmina, V. Sokolovsky, A. Monasova, A. Vorobyov, D. Apeshin. Study of SEIRD adaptive-compartmental model of COVID-19 epidemic spread in Russian Federation using optimization methods // Mathematical biology and bioinformatics, 2021. V. 16. No. 1. Pp. 136-151.
6. O. Krivorotko, S. Kabanikhin. Mathematical models of COVID-19 spread. 2022. URL: <https://arxiv.org/abs/2112.05315>