

АВТОМАТИЧЕСКИЙ ПОДБОР ПАРАМЕТРОВ МОДЕЛИ ARIMA ДЛЯ ПРОГНОЗА КОЛИЧЕСТВА СЛУЧАЕВ ЗАРАЖЕНИЯ И СМЕРТИ ОТ COVID-19

© 2021 Т.А. Макаровских, М.С.А. Аботалеб

Южно-Уральский государственный университет

(454080 Челябинск, пр. им. В.И. Ленина, д. 76)

E-mail: Makarovskikh.T.A@susu.ru, abotalebmostafa@yandex.ru

Поступила в редакцию: 23.11.2020

В работе исследовано применение модели ARIMA прогнозирования временных рядов для анализа открытых данных о распространении коронавирусной инфекции в ряде регионов Российской Федерации. Рассмотрена возможность применения существующих методов и алгоритмов языка программирования для статистической обработки данных R, приводятся алгоритмы подбора параметров модели ARIMA. Разработан и опубликован скрипт на языке программирования R, позволяющий осуществить с помощью стандартной библиотеки `auto.arima` прогнозирование суммарных случаев заражения и летальных исходов на выбранный промежуток времени. В работе показано, что параметры модели различны для временных рядов разной длины, для различных регионов, кроме того, параметры модели меняются с течением времени. Исследован имеющийся инструментарий языка R и показано, что существуют наборы данных, для которых он не позволяет получить параметры модели, дающие наименьшую погрешность. Исследована частота переобучения модели, приведены данные об изменении параметров модели для временных рядов разной длины. Изучение случаев ошибки автоматического подбора параметров модели является темой для дальнейших исследований. Приведена содержательная интерпретация полученных данных. Проведено сравнение прогнозов, полученных в конце октября 2020 г. и актуальных данных на середину ноября 2020 г. Показано, что полученный прогноз позволил достаточно точно предсказать суммарное число заражений и летальных исходов на 7–10 дней.

Ключевые слова: ARIMA, Covid-19, прогнозирование, скрипт, подбор параметров.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Макаровских Т.А., Аботалеб М.С.А. Автоматический подбор параметров модели ARIMA для прогноза количества случаев заражения и смерти от Covid-19 // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2021. Т. 10, № 2. С. 20–37. DOI: 10.14529/cmse210202.

Введение

В конце 2019 года было сообщено о заражениях в провинции Хубэй (Китай) коронавирусной инфекцией нового типа. За несколько месяцев этот новый вирус вызвал глобальную пандемию коронавирусной болезни (CoViD19), которая может закончиться тяжелым острым респираторным синдромом (SARS-COV-2). Несмотря на существование известных эпидемиологических моделей, ни одна из них не может с достаточной долей точности спрогнозировать ситуацию развития эпидемии Covid-19. Исследованию вируса в настоящее время посвящено достаточно много медицинских работ, некоторые из которых не совсем корректно цитируются новостными ресурсами в погоне за сенсацией. Большинство работ является препринтами, т.е. не прошло рецензирование и проверки.

В самом начале эпидемии ежедневно обновлялась информация по долгосрочным прогнозам развития эпидемии в мире в целом и в каждой стране по отдельности. Например, эксперты Сингапурского университета технологии и дизайна (SUTD) прогнозировали, что

эпидемия в России завершится к 20 июня, а в мире — к 9 декабря, потом несколько раз переносили срок и, в конце концов, аннулировали прогнозы. Тем не менее, спустя некоторое время выяснилось, что временные ряды с данными по распространению коронавирусной инфекции пока слишком коротки для составления какого-либо адекватного долгосрочного прогноза. Построение краткосрочных прогнозов высокой точности для распространения подтвержденных случаев заболевания, а также анализа количества смертей и выздоровлений является не менее актуальной задачей.

Чтобы модель адекватно отражала истинное положение распространения инфекции, нужна достаточно точная статистика и подробный анализ внешних факторов, которые не качественно, а именно количественно влияют на распространение инфекции. Например, к таким факторам можно отнести индекс самоизоляции, плотность населения, средний возраст, вероятность заражения и др. Следует отметить, что большинство параметров, которые могли бы повысить точность прогноза, не являются очевидными.

Целью данной работы является подробное исследование работы модели ARIMA для построения краткосрочных прогнозов в условиях непредсказуемости поведения исследуемого процесса. В работе рассматривается возможность автоматического подбора параметров модели ARIMA для временных рядов, соответствующих одному и тому же процессу, протекающему в различных условиях (распространение коронавирусной инфекции в различных регионах РФ). Как показывает обзор имеющихся результатов, подобных исследований и разработок для временных рядов по распространению коронавирусной инфекции в настоящий момент найти не удалось.

Статья организована следующим образом. В первом разделе статьи рассмотрен ряд результатов, полученных российскими и зарубежными аналитиками. Во втором и третьем разделах подробно рассмотрена модель прогнозирования ARIMA и существующие методы в языке программирования для статистической обработки данных R. Приводится алгоритм подбора параметров модели ARIMA и описан созданный скрипт на языке R, для которого входными данными является вектор анализируемых параметров. В четвертом разделе исследована частота переобучения модели, приведены данные об изменении параметров модели для временных рядов разной длины. В заключении сформулированы основные полученные результаты и определены направления для дальнейших исследований.

1. Обзор состояния проблемы

В настоящий момент (конец 2020 г.) опубликовано немало работ, посвященных построению прогнозов с горизонтом 3–7 дней [12, 21, 22]. Все эти работы можно подразделить на три категории относительно используемой модели: модели SIR (классические эпидемиологические модели переходов состояний) [6, 14, 20], модели ARIMA (интегрированные модели авторегрессии и скользящего среднего) [17, 18] и модель Хольта—Уинтерса (адаптивного сглаживания) [6, 17, 19]. Одной из достаточно удачных работ можно считать статью [6], опубликованную в России, в которой авторы рассматривают указанные три модели на временном ряде с марта 2020 г. (начало эпидемии в России) по 10 мая 2020 г.

Классическая модель SIR не позволяет обеспечить высокое качество получаемых прогнозов [14, 20]. В работе [14] используется вышедшее не так давно расширение для R `covid19.analytics`, позволяющее получить полную картину по распространению вируса в любой точке земного шара. Автором разработки утверждается, что он проводит исследования на общедоступных данных, взятых из двух основных источников: «Репозиторий

данных о Covid-19 Центра системных наук и инженерии (CSSE) Университета Джона Хопкинса» [10] для данных по всему миру и США, а данные по городу Торонто берет из [11]. Пакет также предоставляет базовые инструменты и функции анализа и визуализации для исследования этих и других наборов данных, структурированных аналогичным образом. Основным недостатком данной разработки на настоящий момент является использование для прогнозирования исключительно классической модели SIR, которая дает очень большую погрешность. Еще в начале пандемии появились статьи, использующие усложненные модели SIR, учитывающие как выбывание инфицированных из-за смертности (модель SIR(D) в КНР) [20], так и инкубационный период заболевания (модель SEIR) для краткосрочного прогнозирования развития эпидемии в различных странах. Предпринимались попытки разработать модель SEIR в предположении, что развитие эпидемии начинается с подгруппы населения, характеризующейся слабым иммунным ответом, и на более поздней стадии продолжается во всем населении. Хотя это предположение кажется правдоподобным, у авторов подобных работ нет прямого подтверждения того, что неоднородность населения играет значительную роль во время распространения эпидемии, а работы до сих пор опубликованы в Интернете в статусе препринтов. Такие модели не принимают во внимание участие многих различных типов клеток, внутриклеточную регуляцию, задержку по времени в распространении вируса и клональную экспансию иммунных клеток, а также некоторые другие важные аспекты.

Данные модели хорошо подходят для более долгосрочного прогнозирования (более 7 дней). К особенностям применения эпидемиологической модели из семейства SIR(D) можно отнести то, что в ней принимаются постоянными значения коэффициентов, отвечающих за вероятность инфицирования, вероятность излечения и вероятность смерти. Однако в действительности повсеместно вырабатываются и практикуются эффективные (и не очень) меры сдерживания эпидемии (карантины, ограничение перемещения людей, масочный режим и пр.), что сказывается на изменении траектории эпидемии и, как следствие, приводит к тому, что коэффициенты подобной модели становятся переменными. В статье [6] авторы переобучают коэффициенты модели по вновь поступившим данным, что оправдано для получения краткосрочных прогнозов (до 10 дней) высокой точности.

Модели, основанные на анализе временных рядов, в частности ARIMA-модели, являются сложно настраиваемыми при проведении полного анализа, однако дают практически всегда хороший результат там, где требуется качественный прогноз на среднесрочный и краткосрочный период [6]. В работах [16, 18], как и во многих других, авторы вручную подбирают параметры для имеющихся на момент публикации временных рядов и проводят расчеты в широко используемых для статистического анализа программах (R, Gretl) для отдельных государств и регионов для фиксированных дат. Особенно много подобных работ можно найти у аналитиков из Индии и Юго-Восточной Азии. В работе [6] также приводится ряд результатов ученых из Юго-Восточной Азии и Европы, полученных подобным образом. В этой работе указываются параметры модели ARIMA, полученные в начале мая для территории России и отмечается, что для разных регионов данные параметры разнятся, более того, авторы показывают, что для разных периодов развития эпидемии данные параметры различны. В работе не приводится алгоритмов, позволяющих определить параметры в автоматическом режиме. Также хотелось бы отметить работу [13], в которой предлагается рассмотреть отношения между государствами, находящимися в одной географической зоне, для прогнозирования распространения вируса. Страны в одном географическом

ческом регионе имеют переменные с похожими значениями (поддающиеся количественной и не поддающейся количественной оценке), которые влияют на распространение вируса. Такой подход в дальнейшем может быть использован при прогнозировании распространения вируса в РФ, поскольку (как показано в разделе 4) в разных регионах имеет место различный характер распространения эпидемии. Постоянное «дообучение» модели ARIMA на разных этапах развития эпидемии является очевидным недостатком ее использования. Несмотря на то, что зачастую требуется менять не только оцениваемые параметры модели, но и число гиперпараметров модели, данный инструмент хорошо подходит для краткосрочного прогнозирования (сроком до 7 дней).

Достаточно популярным инструментом прогнозирования развития распространения коронавирусной инфекции являются адаптивные модели экспоненциального сглаживания. Данные модели также послужили достаточно распространенным инструментом для построения прогнозов для временных рядов, соответствующих развитию эпидемии в различных странах [8, 17, 19]. Как и в случае с моделью ARIMA, основным недостатком большинства представленных работ является отсутствие объяснения выбора соответствующей спецификации моделей, а также отсутствие «объяснения» подбора гиперпараметров моделей прогнозирования [6]. Отметим также работу [8] одного из соавторов данной статьи, в которой показано, что модель экспоненциального сглаживания для рассматриваемых временных рядов дает более точные результаты по сравнению с моделью ARIMA. Модель Хольта–Уинтерса фактически никак не объясняет сущность процесса развития эпидемии и исключительно ориентируется на сами данные. Таким образом, данная модель может отметить явление незначительной семидневной цикличности, связанной в первую очередь не с истинным развитием процесса инфицирования, а с графиком работы отдельных служб здравоохранения (лабораторий, производящих тестирование, а также административных служб) [6].

Таким образом, получить результаты высокой точности можно только используя динамическое оценивание изменяющихся коэффициентов. Поэтому решение данной задачи является актуальным. Проведенный анализ показывает, что большинство работ по прогнозированию случаев заражения новой коронавирусной инфекцией посвящено исследованию ручного подбора параметров модели ARIMA для данных по одному или нескольким фиксированным регионам; опубликовано очень мало работ (например, [9]), в которых предлагается использовать современные программные средства для прогнозирования временных рядов и в которых описывались бы библиотеки для анализа временных рядов указанного содержания; не удалось найти работ, посвященных особенностям прогнозирования и подбора параметров моделей для регионов Российской Федерации.

2. Алгоритм использования модели ARIMA для прогнозирования временных рядов

В качестве примеров будем использовать временные ряды по заболеваемости Covid-19 в Москве, Челябинской области и России [5] по данным с 12 марта по 27 октября 2020 г. В качестве входных данных предоставляется одномерная последовательность значений. Данные временного ряда представляют собой отсчеты значений наблюдаемых параметров в последовательные дискретные моменты времени.

Модель ARIMA состоит из трех компонентов:

- AR (авторегрессионный термин) — относится к прошлым значениям, используемым для прогнозирования следующего значения; определяется параметром p в модели авторегрессии

$$Y_t = \Theta_1 \cdot Y_{t-1} + \Theta_2 \cdot Y_{t-2} + \dots + \Theta_p \cdot Y_{t-p} + \varepsilon_t,$$

который определяется по PACF (partial auto-correlation function) — «частная корреляционная функция» между Y_t и Y_{t-k} при исключении влияния $Y_{t-1}, \dots, Y_{t-k+1}$.

- MA (скользящее среднее) — используется для определения количества прошлых ошибок прогноза, используемых для прогнозирования будущих значений; определяется параметром q , получаемого из ACF (auto-correlation function, автокорреляционная функция)

$$\rho_k = \frac{\text{cov}\{Y_t, Y_{t-k}\}_t}{\text{var}\{Y_t\}_t},$$

$$y_t = \varepsilon_t + \alpha \cdot \varepsilon_{t-1} + \dots + \alpha_q \cdot \varepsilon_{1-q},$$

где ε_t — белый шум, всегда являющийся стационарным процессом¹. Скользящее среднее показывает наличие колебаний в ряду. Чем выше значение скользящего среднего, тем выше вероятность колебаний.

- I (интегрирующий член) — если ряд не является стационарным, то находится его разность порядка d , которая является стационарным рядом. Для проверки стационарности ряда используются расширенный тест Дики—Фулера (ADF)² тест Квятковского—Филлипса—Шмидта—Шина (Kwiatkowski—Phillips—Schmidt—Shin, KPSS)³. Эти же тесты позволяют определить параметр d модели.

Приведем алгоритм построения модели ARIMA.

Алгоритм ARIMA-Construct

Шаг 1. Загрузить входные данные в вектор TS.

Шаг 2. Проверить, является ли временной ряд TS стационарным. Если TS является стационарным, перейти на шаг 4. Если ряд не является стационарным, то находится его разность порядка d , которая является стационарным рядом.

Шаг 3. Положить $d = 0$.

Шаг 4. Построить графики ACF и PACF для определения входных параметров модели ARIMA.

Шаг 5. С помощью графиков ACF и PACF определить значения p и q для модели ARIMA.

Шаг 6. Выполнить подгонку параметров p и q для идентификации модели ARIMA.

Шаг 7. Спрогнозировать значения на проверочном наборе TS: будущие значения.

Шаг 8. Рассчитать среднеквадратическое отклонение для сравнения прогнозов и фактических значений.

Конец алгоритма

Рассмотрим применение данного алгоритма на временном ряде, соответствующем количеству заражений Covid-19 в день по Челябинской области на конец октября 2020 года. В проведенном эксперименте параметры модели ARIMA подбирались с помощью скриптов

¹Белым шумом называется процесс, имеющий постоянное математическое ожидание, постоянную дисперсию и нулевую, для всех, кроме нулевого лага, автоковариационную функцию.

²Методика, которая используется в прикладной статистике и эконометрике для анализа временных рядов для проверки на стационарность.

³Критерий, используемый для проверки на стационарность наблюдаемого временного ряда.

на языке R. Пример подбора параметров для Челябинской области приведен в [1]. В таблицах приведена описательная статистика по зарегистрированным случаям в день (табл. 1) и летальным случаям (табл. 2). Термины, используемые в таблице, являются общеупотребительными в статистике [7].

Таблица 1

Описательная статистика по заражениям Covid-19 в некоторых регионах РФ

Регион	Среднее	Min	Max	Станд. откл.	Асимм.	Экссесс
Челябинская обл.	7720.7740	0	18 393	6 300.5310	0.0926	1.4626
г. Москва	18 8832.7000	21	405 352	111 203.5000	-0.4602	2.1716
Московская обл.	45 981.6900	4	84 247	27 874.2200	-0.5690	1.7950
г. Санкт-Петербург	23 544.6700	1	57 515	16 584.0400	-0.0078	1.8397
Свердловская обл.	14 259.9300	0	34 782	11 672.7100	0.1144	1.5033
Нижегородская обл.	17 059.3400	1	39 636	12 480.2500	-0.0348	1.6364
Респ. Крым	1 643.7910	0	8 533	2 069.1200	1.5772	4.6731

Таблица 2

Описательная статистика по летальным случаям Covid-19 в некоторых регионах РФ

Регион	Среднее	Min	Max	Станд. откл.	Асимметрия	Экссесс
Челябинская обл.	65.1609	0	159	53.3066	0.0030	1.5226
г. Москва	3 180.0090	0	6503	2 054.9040	-0.4150	1.6976
Московская обл.	752.5633	0	1549	524.2495	-0.2676	1.5088
г. Санкт-Петербург	1 359.0350	0	3627	1 224.3110	0.2248	1.5364
Свердловская обл.	229.2533	0	743	245.7027	0.6538	1.9212
Нижегородская обл.	271.2882	0	721	228.9393	0.2134	1.6182
Респ. Крым	22.6044	0	122	30.8129	1.7321	5.1280

Несмотря на то, что ARIMA является очень мощной моделью для прогнозирования временных рядов, процессы подготовки данных и настройки параметров в итоге отнимают немало времени: требуется сделать ряд стационарным, определить значения p и q с помощью графиков. Кроме того, как было отмечено выше, при анализе данных по пандемии Covid-19 целесообразно использовать динамические параметры моделей, что требует постоянного переобучения модели для каждого отдельно взятого набора данных. Данный факт является причиной появления большого числа исследований, посвященных подбору параметров модели для временных рядов, разнообразных по продолжительности и происхождению. Вопрос о частоте переобучения модели ARIMA для временных рядов по Covid-19, соответствующих различным регионам земного шара, является открытой задачей.

3. Использование функции `auto.arima` в языке R для автоматического подбора параметров модели

Поскольку процесс подбора параметров модели ARIMA является достаточно трудоемким и требует анализа графиков, целесообразно использовать модель AutoARIMA, которая

автоматизирует выполнение шагов 2–5 алгоритма ARIMA-Construct. Приведем алгоритм, использующий автоматический подбор параметров.

Алгоритм AutoARIMA

Шаг 1. Загрузить входные данные (временной ряд) в вектор TS.

Шаг 2. FitAutoARIMA(TS). // Подогнать модель для временного ряда TS.

Шаг 3. Предсказать значения для проверочного набора.

Шаг 4. Проверить производительность модели, используя прогнозные значения в сравнении с фактическими.

Конец алгоритма

Легко видеть, что данный алгоритм полностью обходит выбор функций p и q .

Для анализа временных рядов в R может быть полезна функция `auto.arima`, которая в качестве входных данных принимает одномерные временные ряды и использует алгоритм, который объединяет тесты для подбора лучшей модели ARIMA. Очевидно, что поскольку `auto.arima` — автоматизированная процедура, то ее использование связано с рядом предположений, проверок и противопоставлений. На практике рекомендуется проверка полученных параметров автоматического процесса на соответствие заданным критериям, при необходимости разрешается изменить их. Рассмотрим работу функции `auto.arima` для временных рядов, предложенных в предыдущем разделе. Получим следующие результаты, приведенные в [3, 4]. Таким образом, процесс прогнозирования можно представить в виде следующей структурной схемы (см. рис. 1). Здесь левая ветвь структурной схемы соответствует алгоритму ARIMA-Construct, а правая — алгоритму AutoARIMA. Отметим, что для выполнения всех шагов указанных алгоритмов в R имеются стандартные библиотеки и функции, которые требуется вызвать в необходимой последовательности.

Приведем результаты использования рассмотренного скрипта для временного ряда с данными по суммарному количеству заражений по Челябинской области. Имеем, что по данным до 27 октября включительно лучшая модель для Челябинской области — ARIMA(4,2,0). Графический анализ остатков модели приведен на рис. 2.

Напомним, что остатки в модели временных рядов — это то, что остается после подбора модели. Таким образом, остатки равны разнице между наблюдениями и соответствующими подобранными значениями. Остатки используются для проверки, адекватности реакции модели на информацию о данных. Хороший метод прогнозирования дает остатки со следующими свойствами [15]:

1. Остатки не коррелированы. Если есть корреляции между остатками, значит, в остатках остается информация, которую следует использовать при вычислении прогнозов.
2. Остатки имеют нулевое среднее. Если остатки имеют среднее значение, отличное от нуля, то прогнозы смещены.
3. Остатки имеют постоянную дисперсию.
4. Остатки имеют нормальное распределение.

Рисунок 2 состоит из трех частей, первая часть (верхний график) представляет собой остатки от прогнозирования случаев заражения Covid-19 в Челябинске с использованием наивного метода. Вторая часть (левый нижний рисунок) представляет собой ACF остатков от наивного метода, примененного к случаям инфекции Covid-19. Отсутствие корреляции говорит о том, что получаемые прогнозы имеют достаточно высокое качество. И третья часть (правый нижний график) представляет собой гистограмму остатков от наивного метода, примененного к случаям заражения Covid-19 в Челябинске. Правый хвост представ-

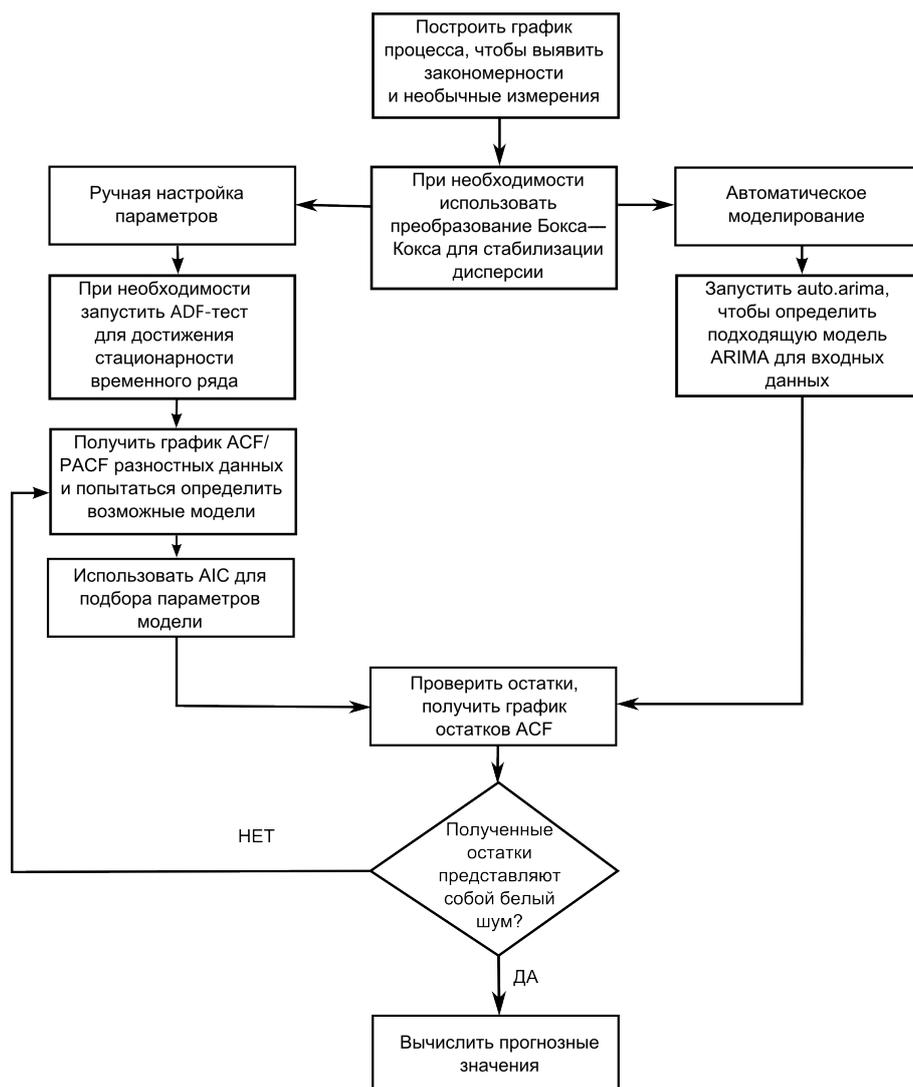


Рис. 1. Структурная схема для алгоритма выбора способа моделирования

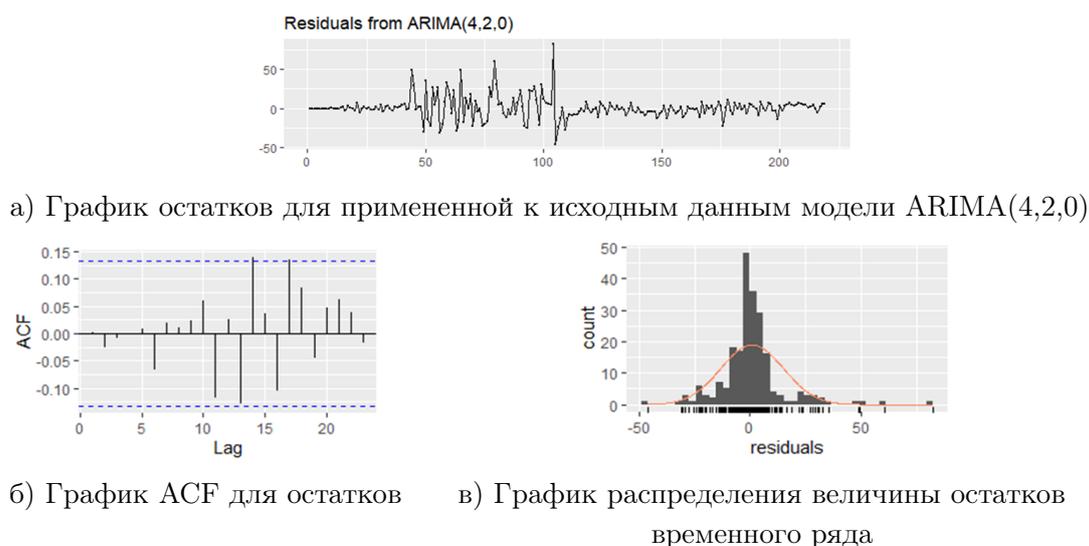


Рис. 2. Графический анализ остатков для модели ARIMA(4,2,0), использованной для прогнозирования числа случаев заражения Covid-19 по Челябинской области

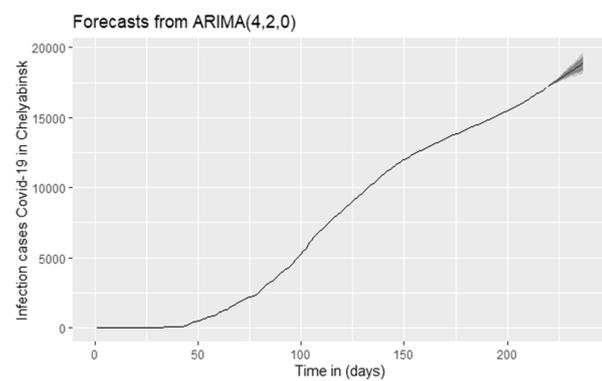


Рис. 3. Прогноз суммарного количества заражений Covid-19 по Челябинской области

ляется слишком длинным для нормального распределения в случаях заражения Covid-19. Эти графики показывают, что наивный метод дает прогнозы, которые, по-видимому, учитывают всю доступную информацию. Среднее значение остатков близко к нулю, и в рядах остатков нет значительной корреляции. Временной график остатков показывает, что вариация остатков остается практически неизменной для исторических данных, за исключением одного выброса, и поэтому остаточную дисперсию можно рассматривать как постоянную. Это также можно увидеть на гистограмме остатков. Гистограмма предполагает, что распределение остатков может не быть нормальным — правый хвост кажется слишком длинным, даже в случае игнорирования выброса. Следовательно, прогнозы этого метода, вероятно, будут довольно хорошими, но интервалы прогнозирования, которые вычисляются с учетом нормального распределения, могут быть неточными. Поэтому проводится тест Льюнга—Бокса⁴, чтобы убедиться, что нет автокорреляции из-за белого шума⁵.

Существует множество методов проверки остатков, удобно упакованных в одну функцию R `checkresiduals()`, которая создаст временной график, график ACF и гистограмму остатков (с наложенным нормальным распределением для сравнения) и проведет тест Льюнга—Бокса с правильными степенями свободы. В рассмотренном случае [3, 4] полученные с помощью программа остатки выглядят как белый шум, что позволяет использовать полученную модель для прогноза. Так, прогноз заболеваемости в Челябинской области на 10 дней вперед приведен на рис. 3. Аналогичные результаты можно получить и для других регионов Российской Федерации.

Разработанный скрипт можно использовать и для прогнозирования случаев смерти в рассмотренных регионах. Отметим, что в этом случае параметры моделей зачастую отличаются от параметров моделей, использованных для прогнозирования суммарного количества заражений (рис. 4). В табл. 3 приведем результаты анализа точности полученных прогнозов. Отметим, что при правильном подборе параметров модели удается с высокой точностью спрогнозировать как суммарное количество случаев заражения, так и суммарное число летальных случаев. Так, при прогнозировании случаев заражений и летальных исхо-

⁴Тест Льюнга—Бокса — статистический критерий, предназначенный для нахождения автокорреляции временных рядов. Вместо тестирования на случайность каждого отдельного коэффициента, он проверяет на отличие от нуля сразу несколько коэффициентов автокорреляции. Обычно тест используется при построении моделей ARIMA. При этом следует иметь в виду, что данное тестирование применяется к остаткам полученной модели, а не к исходным данным.

⁵Белым шумом называется процесс, имеющий постоянное математическое ожидание, постоянную дисперсию и нулевую, для всех, кроме нулевого лага, автоковариационную функцию.

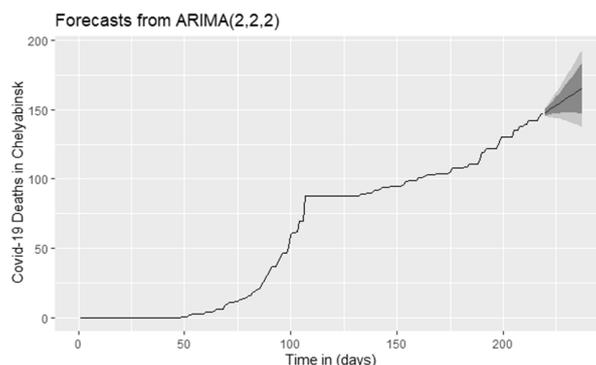


Рис. 4. Полученные лучшие параметры модели ARIMA для Челябинской обл. и результаты прогнозирования суммарного числа летальных случаев Covid-19

Таблица 3

Анализ погрешности прогнозных моделей для Челябинской области

Дата	Случаи заражения			Случаи смерти		
	Факт	Прогноз	Ошибка %	Факт	Прогноз	Ошибка %
17.10.2020	17 223	17 223.78	0.004529	147	147.9774	0.664898
18.10.2020	17 325	17 322.90	0.012121	147	149.2462	1.528027
19.10.2020	17 427	17 420.61	0.036667	152	150.1797	1.197566
20.10.2020	17 526	17 517.87	0.046388	152	151.1371	0.567697
21.10.2020	17 630	17 615.96	0.079637	155	152.1731	1.823806
22.10.2020	17 740	17 714.48	0.143856	155	153.1800	1.174194
23.10.2020	17 857	17 813.21	0.245226	155	154.1754	0.532
24.10.2020	17 986	17 911.82	0.412432	159	155.1815	2.401572
25.10.2020	18 122	18 010.17	0.617095	159	156.1875	1.768868
26.10.2020	18 255	18 108.47	0.802684	159	157.1907	1.137925
27.10.2020	18 393	18 206.77	1.012505	159	158.1948	0.506415
MAPE (%)			0.310285	MAPE (%)		1.209361

дов от Covid-19 по Челябинской области средняя абсолютная процентная ошибка (MAPE) не превышает 1.3%. Здесь и далее в статье в столбце табл. «Прогноз» встречаются вещественные значения, поскольку модель с автоматическим подбором параметров использует непрерывные данные. Округление осуществляется на окончательном этапе при оформлении полученной таблицы и представлении данных для общественности. Для анализа ошибок округлять полученные значения не целесообразно.

Сравним среднюю абсолютную процентную ошибку (MAPE) прогнозов, полученных с помощью `auto.arima`, и прогнозов по данным из работы [6]. Напомним, что модель ARIMA в указанной работе постоянно переобучалась, подбор параметров не осуществлялся автоматически. Будем рассматривать только последний набор данных из работы [6], использующий данные до 10 мая 2020 года включительно. В результате получим следующие значения погрешностей прогноза на неделю (см. табл. 4).

Из таблицы видно, что качество прогнозирования с помощью автоматического подбора параметров сравнимо с использованием моделей с ручным подбором параметров. Отме-

Таблица 4

Сравнение погрешностей прогноза для модели ARIMA с автоматическим подбором параметров и модели, описанной в работе [6] по РФ

Дата	Случаи заражения (модель auto.arima)			Случаи заражения (модель ARIMA, параметры из [6])		
	Факт	Прогноз	Ошибка (%)	Факт	Прогноз	Ошибка (%)
11.05.2020	221 344	220 613.6	0.33	221 344	219 615	0.78
12.05.2020	232 243	231 632.9	0.263	232 243	230 661	0.68
13.05.2020	242 271	242 652.2	0.157	242 271	241 780	0.20
14.05.2020	252 245	253 671.5	0.566	252 245	252 946	0.28
15.05.2020	262 843	264 690.8	0.703	262 843	264 194	0.51
16.05.2020	272 043	275 710.2	1.348	272 043	275 541	1.29
17.05.2020	281 752	286 729.5	1.767	281 752	286 958	1.85
		MAPE (%)	0.733		MAPE (%)	0.799

тим, что получить высокую точность прогнозов возможно только в случае рассмотрения каждого региона по отдельности. Подобрать наиболее подходящую модель для прогнозирования указанных прогнозов для всей страны в целом с такой же высокой точностью оказалось невозможным. Причиной этому является зависимость скорости распространения инфекции от плотности населения, индекса самоизоляции, соблюдения мер безопасности и прочих факторов, значительно различающихся для каждого региона.

Кроме того, несмотря на заметную экономию времени для подбора параметров открытым остается вопрос о частоте переобучения модели ARIMA для анализа временных рядов по Covid-19. Чтобы получить высокий уровень точности и снизить частоту ошибок, лучше всего выполнять этот анализ еженедельно, потому что некоторые изменения, которые менее заметны в случаях заражения и смерти, могут быть реакцией на предпринятые государством либо регионом меры по сдерживанию распространения инфекции. Прочие факторы также могут привести к снижению уровня точности и увеличению ошибки, поэтому целесообразно осуществлять прогнозирование не чаще раза в неделю.

4. Частота переобучения модели ARIMA при анализе данных по Covid-19

В данной работе использованы наборы данных [10, 11], а также официальные данные по регионам РФ, предоставляемые на портале [5]. Одним из результатов работы [6] является настройка моделей, позволяющая (на май 2020 года) получать краткосрочные прогнозы достаточно высокой точности (ошибка порядка 0.3 %). Так, в период 11–17 мая 2020 года средняя абсолютная процентная ошибка по модели Хольта–Уинтерса 0.84 %, по модели SIR(D), дающей самый пессимистичный прогноз, — 1.1 %, по модели ARIMA — 0.8 %. Будем в дальнейшем использовать данные значения как ориентир.

Рассмотрим подбор параметров модели ARIMA при загрузке в разработанный скрипт входного вектора с данными на последний день каждого месяца (с апреля по октябрь) для рассматриваемых регионов (табл. 5).

Отметим, что используется стандартное кодирование ARIMA (p, d, q), где параметры заменяются целочисленными значениями. Каждый из этих компонентов явно указан в мо-

Таблица 5

Параметры моделей ARIMA при загрузке временного ряда на конец каждого месяца
(с апреля по октябрь 2020 г.)

Регион		30.04	31.05	30.06	31.07	31.08	30.09	31.10
Челябинская обл.	Заражения	(0,2,0)	(2,2,3)	(4,2,0)	(4,2,0)	(4,2,0)	(4,2,0)	(4,2,0)
	Смерти	(0,0,0)	(0,0,1)	(4,2,0)	(1,2,3)	(1,2,3)	(1,2,3)	(1,2,3)*
г. Москва	Заражения	(0,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(1,2,0)	(2,2,2)	(3,2,2)*
	Смерти	(2,2,0)	(0,2,2)	(0,2,2)	(1,2,1)	(1,2,1)	(1,2,1)	(2,2,2)*
Респ. Крым	Заражения	(0,2,1)	(0,2,1)	(0,2,5)	(2,2,3)	(0,2,5)	(0,2,5)	(5,2,0)*
	Смерти	(0,0,0)	(0,1,2)	(0,2,1)	(0,2,1)	(0,2,1)	(2,2,2)	(3,2,1)*
Московская обл.	Заражения	(0,2,1)	(1,2,0)	(1,2,0)	(1,2,2)	(1,2,2)	(1,2,2)	(1,2,2)
	Смерти	(0,2,5)	(2,2,3)	(0,2,5)	(0,2,5)	(0,2,5)	(0,2,5)	(0,2,5)
г. С.Петербург	Заражения	(0,2,1)	(4,2,0)	(4,2,0)	(4,2,0)	(0,2,5)	(0,2,5)	(0,2,5)
	Смерти	(0,2,4)	(0,2,1)	(0,2,3)	(4,2,0)	(1,2,3)	(1,2,3)	(1,2,3)*
Свердловская обл.	Заражения	(1,2,0)	(0,2,2)	(0,2,1)	(0,2,1)	(0,2,1)	(2,2,3)	(2,2,3)
	Смерти	(0,1,0)	(0,2,1)	(5,2,0)	(2,2,3)	(2,2,3)	(2,2,3)	(1,2,2)
Нижегородская обл.	Заражения	(0,2,1)	(1,2,3)	(5,2,0)	(0,2,5)	(0,2,5)	(0,2,5)	(0,2,5)
	Смерти	(0,2,1)	(1,2,1)	(5,2,0)	(1,2,2)	(2,2,3)	(0,2,1)	(0,2,1)

дели в виде коэффициента. При изучении временных рядов по коронавирусной инфекции в регионах было определено, что для них имеет место стационарность второго порядка (второй параметр для всех полученных моделей равен 2). Принципиальное различие между моделями AR и MA основано на корреляции между объектами временных рядов в разные моменты времени. Ковариация между Y_t и Y_{t-n} равна нулю для моделей MA. Однако корреляция между Y_t и Y_{t-n} постепенно уменьшается по мере увеличения n в модели AR. Это означает, что модель скользящего среднего (MA) не использует предыдущие прогнозы для прогнозирования будущих значений при использовании ошибок из предыдущих прогнозов. В то время как модель авторегрессии (AR) использует предыдущие прогнозы для прогнозирования будущих значений. Данный метод позволяет узнать, информацию по скольким дням из прошлого необходимо использовать для получения прогноза будущих значений. Параметр определяется путем вычисления автокорреляции между количеством случаев в настоящий момент и по прошедшим дням, пока не будет достигнуто максимальное количество дней в прошлом, которые модель ARIMA может использовать с наибольшим весом. Например, AR (3) означает, что предсказанный новый день зависит от трех предшествующих ему дней. Так, модель ARIMA Санкт-Петербурга для случаев (0,2,5) 27 октября 2020 г. и степень MA равна 5, что означает, что в октябре есть колебания в количестве выявленных случаев.

Заметим, что в последнем столбце табл. 5 ряд значений отмечено звездочкой, это данные, для которых с помощью модели `auto.arima` не удалось получить наименьшую погрешность прогнозирования. Из табл. 5 видно, что имеет место изменение параметров модели с течением времени и накоплением дополнительной информации. Вопрос о скорости изменения параметров и «переломных» моментах, в которые имеет место смена параметров прогнозной модели остается открытым и является темой дальнейших исследований.

Рассмотрим интерпретацию полученных результатов для Челябинской области. Аналогичные рассуждения можно провести и для других регионов. Ожидается, что в ближайший период в Челябинске продолжится рост случаев заражения, количество случаев заражения через неделю превысит 20 тысяч, поэтому необходимо принять дополнительные меры для

предотвратить такой рост случаев инфицирования. В ближайшее время ожидается небольшое увеличение ежедневной смертности в Челябинске, но в ноябре 2020 года оно в сумме не превысит 200 случаев. Как следует из имеющейся статистики [5] на 19 ноября 2020 года, количество заражений равно 22627, что соответствует полученному прогнозу. Количество летальных исходов — 193. Таким образом, спрогнозированный предел будет достигнут уже 20 ноября. На дату подачи статьи разработанная модель позволяет сделать следующий прогноз: количество случаев заражения составит более 220 случаев в сутки, а смертность достигнет 15 случаев в сутки. Отметим, что полученные результаты позволяют судить о достаточно хорошем качестве краткосрочных прогнозов с использованием модели ARIMA. Таких прогнозов зачастую может быть достаточно для принятия управленческих решений.

Заключение

Пандемия коронавируса стала глобальным вызовом для науки: не только для медицины, но и для математики и статистики. Требуется разработка новых подходов к моделированию и прогнозированию временных рядов в условиях непредсказуемости поведения исследуемого процесса. Говорить о построении модели, которая адекватно описывала бы процесс распространения коронавирусной инфекции, в настоящий момент преждевременно, поскольку не прошло еще года с начала пандемии, нет возможности оценить возможный сезонный характер данных.

К основным полученным результатам можно отнести следующие.

1. Показано, что параметры модели различны для различных регионов, кроме того, параметры модели меняются с течением времени. Исследование закономерностей изменения параметров модели остается открытой задачей. В работе рассматривается возможность автоматического подбора параметров модели ARIMA для временных рядов, соответствующих одному и тому же процессу, протекающему в различных условиях (распространение коронавирусной инфекции в различных регионах РФ).

2. Разработан скрипт на языке программирования R, позволяющий осуществить с помощью стандартной библиотеки `auto.arima` прогнозирование суммарных случаев заражения и летальных исходов на выбранный промежуток времени в указанном пользователем регионе или государстве. Показано, что автоматический подбор параметров в ряде случаев не позволяет выбрать, действительно, лучшую модель. Исследование случаев ошибки автоматического подбора параметров модели является темой для дальнейших исследований. Разработанный скрипт включен в общую библиотеку `epidemic.ta` скриптов для прогнозирования временных рядов по Covid-19.

В качестве направлений дальнейших исследований можно выделить изучение и выявление особенностей использования прочих классических моделей прогнозирования временных рядов, пополнение библиотеки `epidemic.ta` [2], позволяющей получать и исследовать прогнозы с помощью различных методов, а также рассмотрение применимости неклассических (квазилинейных) моделей для построения прогнозов развития коронавирусной инфекции и применимости полученных моделей для прогнозирования временных рядов схожей структуры для прочих прикладных задач, например, для задач экономики, логистики, распространения информации в социальных сетях и пр.

Статья выполнена при поддержке Правительства РФ (Постановление № 211 от 16.03.2013 г.), Соглашение № 02.A03.21.0011 и Министерства науки и высшего образования РФ (государственное задание FENU-2020-0022).

Литература

1. Аботалеб М.С.А., Макаровских Т.А. Ручной подбор параметров модели ARIMA для прогнозирования летальных случаев Covid-19 в Челябинский области. URL: <https://rpubs.com/abotalebmostafa/710535> (дата обращения: 09.01.2021).
2. Аботалеб М.С.А., Макаровских Т.А. New package «Epidemic.ta» for forecasting Covid-19 infection cases apply Example For forecasting infection cases in the USA. URL: <https://rpubs.com/abotalebmostafa/711262> (дата обращения: 09.01.2021).
3. Аботалеб М.С.А., Макаровских Т.А. Research opportunities automatic parameter selection arima models for short term forecast cases Covid-19. URL: <https://rpubs.com/abotalebmostafa/710470> (дата обращения: 09.01.2021).
4. Аботалеб М.С.А., Макаровских Т.А. Using auto arima for forecasting Deaths of COVID-19 in Chelyabinsk. URL: <https://rpubs.com/abotalebmostafa/710528> (дата обращения: 09.01.2021).
5. Коронавирус: статистика. URL: <https://yandex.ru/covid19/stat> (дата обращения: 09.01.2021).
6. Лакман И.А., Агапитов А.А., Садикова Л.Ф. и др. Возможности математического прогнозирования коронавирусной инфекции в Российской Федерации // Артериальная гипертензия. 2020. Т. 26, № 3. С. 288–294. DOI: 10.18705/1607-419X-2020-26-3-288-294.
7. Эверитт Б. Большой словарь по статистике. СПб.: Проспект, 2012. 736 с.
8. Abotaleb M.S.A. Predicting COVID-19 Cases using Some Statistical Models: An Application to the Cases Reported in China Italy and USA // Academic Journal of Applied Mathematical Sciences. 2020. Vol. 6, no. 4. P. 32–40. DOI: 10.32861/ajams.64.32.40.
9. Chaurasia V., Pal S. COVID-19 Pandemic: ARIMA and Regression Model-Based Worldwide Death Cases Predictions // SN Computer Science. 2020. Art. 1:288. DOI: 10.1007/s42979-020-00298-6.
10. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: <https://github.com/CSSEGISandData/COVID-19> (дата обращения: 23.11.2020).
11. COVID-19: Status of Cases in Toronto. URL: <https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/> (дата обращения: 23.11.2020).
12. Elie R., Hubert E., Turinici G. Contact rate epidemic control of COVID-19: an equilibrium view // Mathematical Modelling of Natural Phenomena. 2020. Vol. 15, art. 35. DOI: 10.1051/mmnp/2020022.
13. Hernandez-Matamorosb A., Fujitaa H., Hayashib T., Perez-Meana H. Forecasting of COVID-19 per regions using ARIMA models and polynomial functions // Applied Soft Computing Journal. 2020. Vol. 96, art. 106610. DOI: 10.1016/j.asoc.2020.106610.
14. Hussain N., Li B. Using R-studio to examine the COVID-19 Patients in Pakistan Implementation of SIR Model on Cases // International Journal of Scientific Research in Multidisciplinary Studies. 2020. Vol. 6, no. 8. P. 54–59. DOI: 10.13140/RG.2.2.32580.04482.
15. Hyndman R.J., Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018. 384 p.

16. Kumar M., Gupta S., Kumar K., Sachdeva M. Spreading of Covid-19 in India, Italy, Japan, Spain, UK, US: A Prediction Using ARIMA and LSTM Model // Digital Government: Research and Practice. 2020. Vol. 1, no. 4, art. 24. DOI: 10.1145/3411760.
17. Panda M. Application of ARIMA and Holt–Winters forecasting model to predict the spreading of COVID-19 for India and its states. DOI: 10.1101/2020.07.14.20153908 (дата обращения: 10.02.2021).
18. Perone G. ARIMA forecasting of COVID-19 incidence in Italy, Russia, and the USA. DOI: 10.2139/ssrn.3612402 (дата обращения: 10.02.2021).
19. Shokeralla A.A.A., Sameeh F.R.I., Musa A.G.M., Zahrani S. Prediction the daily number of confirmed cases of covid-19 in Sudan with arima and holt winter exponential smoothing // International Journal of Development Research. 2020. Vol. 10, no. 8. P. 39408–39413 DOI: 10.37118/ijdr.19811.08.2020.
20. Sun D., Duan L., Xiong J., Wang D. Modelling and forecasting the spread tendency of the COVID-19 in China // BMC Infectious Diseases. DOI: 10.21203/rs.3.rs-26772/v1 (дата обращения: 10.02.2021).
21. Volpert V., Banerjee M., d’Onofrio A., Lipniacki T., Petrovskii S., Tran V.Ch. Coronavirus: scientific insights and societal aspects // Mathematical Modelling of Natural Phenomena. 2020. Vol. 15, art. E2. DOI: 10.1051/mmnp/2020010.
22. Volpert V., Banerjee M., Petrovskii S., Volpert V. On a quarantine model of coronavirus infection and data analysis // Mathematical Modelling of Natural Phenomena. 2020. Vol. 15, art. 24. DOI: 10.1051/mmnp/2020006.

Макаровских Татьяна Анатольевна, д.ф.-м.н., доцент, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

Аботалёв Мостафа Салахелдин Абделсалам, аспирант, кафедра системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет) (Челябинск, Российская Федерация)

AUTOMATIC SELECTION OF ARIMA MODEL PARAMETERS TO FORECAST COVID-19 INFECTION AND DEATH CASES

© 2021 Т.А. Makarovskikh, М.С.А. Abotaleb

South Ural University (pr. Lenina 76, Chelyabinsk, 454080 Russia)

E-mail: Makarovskikh.T.A@susu.ru, abotalebmostafa@yandex.ru

Received: 23.11.2020

In our paper we explore the use of the ARIMA model for forecasting time series for the analysis of open data on the spread of the coronavirus infection in a number of the Russian Federation regions. The possibility of using the existing methods and algorithms of R programming language is considered, algorithms for selecting the parameters of the ARIMA model are presented. We have developed and uploaded the script in R programming language, which allows using the standard library `auto.arima` to predict the total cases of infection and deaths for a selected period. The paper shows that the parameters of the model are different for time series of different lengths, for different regions; in addition, the parameters of the model change over time. The available toolkit of the R language is investigated and it is shown that there are data sets for which it does not allow obtaining the parameters of the model that gives the smallest error. The frequency of model retraining is investigated, data on changes in the model parameters for time series of different lengths are presented. Investigation of cases of errors in automatic selection of model parameters is a topic for further research. We have presented a meaningful interpretation of the data obtained. A comparison of the forecasts obtained at the end of October, 2020 and actual data for the middle of November, 2020 is carried out. We have shown that the obtained forecast made it possible to accurately predict the total number of infections and deaths for 7–10 days for any further period.

Keywords: ARIMA, Covid-19, forecasting, script, parameters selection.

FOR CITATION

Makarovskikh T.A., Abotaleb M.S.A. Automatic Selection of ARIMA Model Parameters to Forecast COVID-19 Infection and Death Cases. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2021. Vol. 10, no. 2. P. 20–37. (in Russian) DOI: 10.14529/cmse210202.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Abotaleb M.S.A., Makarovskikh T.A. Manual selection of ARIMA model parameters for predicting death cases of Covid-19 in the Chelyabinsk region. URL: <https://rpubs.com/abotalebmostafa/710535> (accessed: 09.01.2021).
2. Abotaleb M.S.A., Makarovskikh T.A. New package “Epidemic.ta” for forecasting Covid-19 infection cases apply Example For forecasting infection cases in the USA. URL: <https://rpubs.com/abotalebmostafa/711262> (accessed: 09.01.2021).
3. Abotaleb M.S.A., Makarovskikh T.A. Research opportunities automatic parameter selection ARIMA models for short term forecast cases Covid-19. URL: <https://rpubs.com/abotalebmostafa/710470> (accessed: 09.01.2021).
4. Abotaleb M.S.A., Makarovskikh T.A. Using auto ARIMA for forecasting Deaths of COVID-19

- in Chelyabinsk. URL: <https://rpubs.com/abotalebmostafa/710528> (accessed: 09.01.2021).
5. Coronavirus: statistics. URL: <https://yandex.ru/covid19/stat> (accessed: 09.01.2021).
 6. Lakman I.A., Agapitov A.A., Sadikova L.F. et al. Possibilities of mathematical forecasting of coronavirus infection in the Russian Federation. *Arterialnaya gipertenzia*. 2020. Vol. 26, no. 3. P. 288–294. (in Russian) DOI: 10.18705/1607-419X-2020-26-3-288-294.
 7. Everitt B.S. *The Cambridge Dictionary of Statistics*. The Cambridge Dictionary of Statistics. Ed. 4. 2010. 736 p.
 8. Abotaleb M.S.A. Predicting COVID-19 Cases using Some Statistical Models: An Application to the Cases Reported in China Italy and USA. *Academic Journal of Applied Mathematical Sciences*. 2020. Vol. 6, no. 4. P. 32–40. DOI: 10.32861/ajams.64.32.40.
 9. Chaurasia V., Pal S. COVID-19 Pandemic: ARIMA and Regression Model-Based Worldwide Death Cases Predictions. *SN Computer Science*, 2020. Art. 1:288. DOI: 10.1007/s42979-020-00298-6.
 10. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: <https://github.com/CSSEGISandData/COVID-19> (accessed: 23.11.2020).
 11. COVID-19: Status of Cases in Toronto. URL: <https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/> (accessed: 23.11.2020).
 12. Elie R., Hubert E., Turinici G. Contact rate epidemic control of COVID-19: an equilibrium view. *Mathematical Modelling of Natural Phenomena*. 2020. Vol. 15, art. 35. DOI: 10.1051/mmnp/2020022.
 13. Hernandez-Matamorosb A., Fujitaa H., Hayashib T., Perez-Meana H. Forecasting of COVID-19 per regions using ARIMA models and polynomial functions. *Applied Soft Computing Journal*. 2020. Vol. 96, art. 106610. DOI: 10.1016/j.asoc.2020.106610.
 14. Hussain N., Li B. Using R-studio to examine the COVID-19 Patients in Pakistan Implementation of SIR Model on Cases. *International Journal of Scientific Research in Multidisciplinary Studies*. 2020. Vol. 6, no. 8. P. 54–59. DOI: 10.13140/RG.2.2.32580.04482.
 15. Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. OTexts, 2018. 384 p.
 16. Kumar M., Gupta S., Kumar K., Sachdeva M. Spreading of Covid-19 in India, Italy, Japan, Spain, UK, US: A Prediction Using ARIMA and LSTM Model. *Digital Government: Research and Practice*. 2020. Vol. 1, no. 4, art. 24. DOI: 10.1145/3411760.
 17. Panda M. Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. DOI: 10.1101/2020.07.14.20153908 (accessed: 10.02.2021).
 18. Perone G. ARIMA forecasting of COVID-19 incidence in Italy, Russia, and the USA. DOI: 10.2139/ssrn.3612402.
 19. Shokeralla A.A.A., Sameeh F.R.I., Musa A.G.M., Zahrani S. Prediction the daily number of confirmed cases of Covid-19 in Sudan with ARIMA and Holt-Winter exponential smoothing. *International Journal of Development Research*. 2020. Vol. 10, no. 8. P. 39408–39413 DOI: 10.37118/ijdr.19811.08.2020.

20. Sun D., Duan L., Xiong J., Wang D. Modelling and forecasting the spread tendency of the COVID-19 in China. BMC Infectious Diseases. DOI: 10.21203/rs.3.rs-26772/v1.
21. Volpert V., Banerjee M., d'Onofrio A., Lipniacki T., Petrovskii S., Tran V.Ch. Coronavirus: scientific insights and societal aspects. Mathematical Modelling of Natural Phenomena. 2020. Vol. 15, art. E2. DOI: 10.1051/mmnp/2020010.
22. Volpert V., Banerjee M., Petrovskii S., Volpert V. On a quarantine model of coronavirus infection and data analysis. Mathematical Modelling of Natural Phenomena. 2020. Vol. 15, art. 24. DOI: 10.1051/mmnp/2020006.