



УДК 004.94

ОБЗОР И АНАЛИЗ МАТЕМАТИЧЕСКИХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ЭПИДЕМИОЛОГИЧЕСКОГО ПРОЦЕССА НА ПРИМЕРЕ ГРИППА И ОРВИ

*Д.В. Гринченков, grindv@yandex.ru, Д.Н. Куций, dkushchiy@rambler.ru,
И.А. Спиридонова, sia1706@yandex.ru, Н.И. Химичев, himichevni@mail.ru*
Южно-Российский государственный политехнический университет (НПИ)
имени М.И. Платова, г. Новочеркасск

В статье рассматривается задача прогнозирования эпидемиологического процесса в условиях современного уровня информатизации. Представлен обзор и выполнен анализ современных методов, используемых для формирования краткосрочных и долгосрочных прогнозов. Рассмотрены основные категории формальных методов: анализ временных рядов, машинное обучение и решение задачи фильтрации. Приведены основные направления по разработке математических моделей распространения заболеваний. Выделены достоинства и ограничения моделей и методов в практическом применении. В статье приводятся данные вирусологического мониторинга за гриппом и ОРВИ, проводимого Управлением Роспотребнадзора по Ростовской области.

Ключевые слова: математическая эпидемиология, прогнозирование заболеваемости, методы прогнозирования

REVIEW AND ANALYSIS OF MATHEMATICAL PREDICTION METHODS FOR THE EPIDEMIOLOGICAL PROCESS ON THE EXAMPLE OF INFLUENZA AND ARVI

D.V. Grinchenkov, D.N. Kushchiy, I.A. Spiridonova, N.I. Himichev
Platov South-Russian State Polytechnic University (NPI), Novocherkassk

In article the problem of forecasting of epidemiological process in the conditions of the modern level of informatization is considered. The review is submitted and the analysis of the modern methods used for formation of short-term and long-term forecasts is made. The main categories of formal methods are considered: analysis of temporary ranks, machine learning and solution of a problem of filtration. The main directions on development of mathematical models of spread of diseases are given. Advantages and restrictions of models and methods in practical application are allocated. The data of virological monitoring of influenza and ARVI conducted by Federal Service for Consumer Protection and Human Welfare in the Rostov region are used in article.

Keywords: mathematical epidemiology, morbidity forecasting, forecasting methods

На сегодняшний день самыми распространенными инфекционными болезнями в мире являются грипп и ОРВИ. Кроме того, они также характеризуются повсеместным отсутствием тенденции к снижению уровня заболеваемости. Широкое распространение данной группы инфекций обуславливает занимаемое ими первое место по наносимому экономическому ущербу.

Своевременный и адекватный прогноз является необходимым условием для планирования структуры, масштабов, сроков мероприятий, направленных на предупреждение эпидемий и вспышек заболеваемости, а также на сокращение их негативных последствий. Ретроспективный анализ, выполненный в работе [1] позволяет установить, что Ростовская область вовлекается в общую эпидемию в стране с последней декады января или первой недели февраля до середины марта,



по тенденции развития эпидемического процесса выделены периоды: эпидемический (январь – март), послепандемический (март – апрель), межэпидемический (май – октябрь) и предэпидемический (ноябрь – декабрь). Региональными Управлениями Роспотребнадзора собирается информация о числе граждан, обратившихся в медицинские учреждения и получивших диагноз грипп или ОРВИ, и затем она обрабатывается и добавляется на сайт службы каждую неделю (рис. 1), за исключением межэпидемического сезона.



Рис. 1 – Результаты еженедельного мониторинга заболеваемости гриппом и ОРВИ

Обработка полученной информации включает в себя также и расчет эпидемических порогов, исходными данными для вычисления которого являются статистические сведения о суммарной заболеваемости гриппом и ОРВИ в интенсивных показателях на 10 тыс. населения за календарные недели каждого года в течение последних 5-10 лет. Верхний толерантный предел для каждой календарной недели рассчитывается с помощью значения критерия Стьюдента для доверительной вероятности 95% с N_i-2 степенями свободы. Параметр N_i – количество лет из рассматриваемого промежутка времени [2].

На рисунке 2 представлен сезонный график заболеваемости, полученный в результате обработки информации, представленной на сайте Роспотребнадзора по Ростовской области. В тоже время показатели заболеваемости недели за несколько лет формируют временной ряд y_t , который следует рассматривать как одну из возможных реализаций случайного процесса заболеваемости [3].

В целом ряде научных работ рассмотрены различные методы и способы построения математических моделей прогнозирования эпидемических ситуаций, применение которых должно давать лучший результат по сравнению с применяемым усреднением данных [4-7]. На основании проведенного анализа источников можно выделить три основные группы методов (рис. 3), применяемых для прогнозирования инфекционных заболеваний: статическая обработка данных, машинное обучение и фильтрация.



Рис. 2 – Сезонные эпидемические графики гриппа и ОРВИ по Ростовской области



Рис. 3 – Классификация методов прогнозирования заболеваний

Точечная оценка, относящаяся к группе статистических методов, предполагает вычисление по каждому сечению процесса заболеваемости усредненного показателя по неделям:

$$y_t' = \sum_{j=1}^n y_{t-jT} / n,$$

где y_t' – предполагаемый показатель заболеваемости; T – период сезонности; n – количество доступных наблюдений.

Данный метод прост в применении. Однако, в силу отсутствия учета корреляции между последовательными наблюдениями, он может использоваться только для получения грубых среднесрочных и долгосрочных прогнозов заболеваемости.

Регрессионная модель прогнозирования гриппа должна учитывать сезонный характер заболеваемости, что может быть достигнута с помощью применения циклической функции Серфлинга [4]:

$$y_t' = \sum_{j=0}^v \alpha_j t^j + \sum_{j=1}^k (\beta_{2j-1} \sin \Theta_j + \beta_{2j} \cos \Theta_j),$$



где α_j и β_j – параметры регрессии; ν – степень; θ_j – линейная функция времени t , обычно равная $\theta_j = 2\pi jt/T$; k – число гармоник.

Регрессионный анализ включает неадаптивные и адаптивные модели. Первый вид моделей не подходит для краткосрочного прогнозирования. Адаптивные модели, напротив, наиболее эффективны для таких оценок.

Моделирование основывается на предположении, что последующие или предшествующие значения коррелируют с текущим. Причем близко лежащие значения оказывают большее влияние, нежели далеко стоящие. То есть механизм прогнозирования основывается на связи между значениями и том факте, что она сохранится в дальнейшем.

Пусть n – порядок модели, тогда авторегрессионные модели можно записать следующим образом [5]:

$$y'_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_n y_{t-n} + \varepsilon,$$

где $\alpha_1, \dots, \alpha_n$ – коэффициенты авторегрессии; y_{t-n} – показатель заболеваемости на несколько периодов ранее; ε – случайная ошибка соответствия модели наблюдениям.

Динамические байесовские сети – категория методов, относящихся к прогнозированию на основе машинного обучения, – позволяют учитывать структуру корреляции данных. Байесовские сети изображают в форме направленного графа, вершины которого соответствуют переменным модели, а ребра – вероятностным зависимостям между ними, которые заданы определенными законами распределения. После обучения байесовские сети позволяют оценить вероятность наступления некоторого события при наблюдаемой последовательности явлений. Байесовские сети применяются к задаче прогнозирования заболеваемости преимущественно в простой форме скрытых марковских моделей.

Подход к прогнозированию с использованием искусственных нейронных сетей является громоздким, непрозрачным и достаточно сложным. Недостатком является то, что у разработчика есть возможность формировать входы и наблюдать выходы, но нет возможности проследить, как рассчитываются эти значения. То есть отсутствует доступ к внутренней структуре сети. Такие модели требуют большой выборки данных, однако обеспечивают высокую эффективность прогнозирования.

Рассуждения на основе прецедентов – семейство методов, базирующееся на идее поиска ответа на поставленную задачу среди уже известных способов ее решения. Применительно к задаче прогнозирования заболеваемости этот подход называют методом аналогий. Он сводится к поиску среди накопленной истории наблюдений эпидемических ситуаций близких значений к сложившемуся уровню на настоящий момент времени, и построению прогноза, исходя из последствий этих ситуаций.

В этом случае прогноз может быть получен по формуле [6]:

$$y'_t = \sum_{i=1}^n w_j y_{tj},$$

где y_{t1}, \dots, y_{tm} – показатели заболеваемости прошлых лет, аналогичные текущей ситуации; w_j – весовой коэффициент близости значений.

Любые временные ряды заболеваемости можно рассматривать как случайный процесс, состоящий из сигнала, отражающего реальную эпидемическую об-



становку, и высокочастотного шума. Фильтрация шума позволяет уточнить прогноз и может выполняться как в ходе предварительной обработки исходных данных, так и в составе самого алгоритма прогнозирования.

Отфильтровать измерения можно многими способами, один из них – вейвлет-декомпозиция. В ее основе лежит разложение временного ряда с помощью вейвлет-функций ϕ и ψ [7]:

$$y'_t = \sum_{k=1}^n a_k \phi(t-k) + \sum_{k=1}^n \sum_{j=1}^m d_{j,k} \Psi(2^j t - k),$$

где a_k – отмасштабированный сигнал в момент времени k ; $d_{j,k}$ – детализирующий коэффициент; n и m – целые числа.

Фильтрация достигается за счет ограничения множества значений коэффициентов разложения.

К методам фильтрации относится также экспоненциальное сглаживание. При экспоненциальном сглаживании значение заболеваемости y_t в момент времени t описывается взвешенной суммой последних наблюдений [5]:

$$U_{t+1} = \alpha y_t + (1-\alpha)U_t,$$

где U_t – экспоненциально взвешенная средняя для ряда; α – параметр сглаживания.

Прогноз при этом рассчитывается как $y'_t = U_{t+1}$, где U_{t+1} – сглаженное значение. Данный метод наиболее эффективен для получения среднесрочных прогнозов.

Все вышеперассмотренные методы можно отнести к общим методам анализа временных рядов. В том или ином виде они могут применяться при прогнозировании данных любой природы.

Одним из наиболее простых методов в программной реализации, за исключением точечных оценок, является авторегрессионный анализ, модель которого ARIMA (Autoregressive Integrated Moving Average) поддерживается базовой библиотекой моделирования pandas языка Python [8].

Список цитируемой литературы

1. Ковалев Е.В., Слись С.С., Ненадская С.А., Мирошниченко Г.А., Крат А.В., Баташева И.И., Говорухина М.В., Литовко А.Р. Опыт проведения организационно-профилактических мероприятий в Ростовской области в эпидсезон гриппа 2015 - 2016 годов. Эпидемиология и Вакцинопрофилактика. 2017;16(2):88-95. <https://doi.org/10.31631/2073-3046-2017-16-2-88-95>.
2. ОМР 3.1.2.0102-15 Методика расчета эпидемических порогов по гриппу и острым респираторным вирусным инфекциям по субъектам Российской Федерации. Доп. 2 к МР 3.1.2.0005-10 – Режим доступа https://www.rosпотребнадзор.ru/documents/details.php?ELEMENT_ID=6288 (дата обращения: 10.03.2019).
3. Леоненко В.Н., Новоселова Ю.К., Онг К.М. Предсказание пиков эпидемий гриппа в Санкт-Петербурге с помощью популяционных математических моделей // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 6. С. 1145–1148. doi: 10.17586/2226-1494-2016-16-6-1145-1148.
4. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование, 2013, т. 5, № 5, с. 863-882
5. Цапина Д.Д., Паскевич А.А. Анализ и предварительная обработка данных для решения задач прогнозирования в сфере услуг здравоохранения [Электронный ресурс] // Молодой ученый. – 2017. – №27. – С. 15-24. – Режим доступа <https://moluch.ru/archive/161/45009/> (дата обращения: 10.03.2019).



6. Википедия [Электронный ресурс] – Режим доступа: https://ru.wikipedia.org/wiki/Рассуждения_на_основе_прецедентов – статья «Рассуждение на основе прецедентов» (дата обращения 15.03.2019).
7. Авилов К.К., Романюха А.А. Математические модели распространения и контроля туберкулеза [Электронный ресурс] // Математическая биология и биоинформатика. – 2007. – Т. 2, № 2. – С. 188 – 318. – Режим доступа: [http://www.matbio.org/downloads/Avilov2007\(2_188\).pdf](http://www.matbio.org/downloads/Avilov2007(2_188).pdf) (дата обращения: 15.03.2019).
8. 8HOST [Электронный ресурс] – Режим доступа: <https://www.8host.com/blog/prognozirovanie-vremennyyh-ryadov-s-pomoshhyu-arima-v-python-3/> – статья «Прогнозирование временных рядов с помощью ARIMA в Python3» (дата обращения 21.04.2019).

© Д.В. Гринченков, Д.Н. Куций, И.А. Спиридонова, Н.И. Химичев, 2019