

В.И. ВЬЮН, Т.К. ЕРЕМЕНКО, Г.Е. КУЗЬМЕНКО, Ю.А. МИХНЕНКО

ОБ ОДНОМ ПОДХОДЕ К ПРОГНОЗИРОВАНИЮ ЭПИДЕМИОЛОГИЧЕСКОЙ ОБСТАНОВКИ ПО ГРИППУ-ОРВИ С ИСПОЛЬЗОВАНИЕМ ВРЕМЕННЫХ РЯДОВ

Анотація. Автори досліджують можливість застосування методології та аналітики багатовимірних часових рядів для короткострокового прогнозування епідеміологічного процесу. Для моделювання використовувалися реальні дані моніторингу епідемії Грипп-ГРВІ в м.Києві (офіційний сайт МОЗ за листопад 2009 – січень 2010).

Ключові слова: короткострокове прогнозування, епідеміологічні процеси, тимчасові ряди, моніторинг, моделювання.

Аннотация. Авторы исследуют возможность применения методологии и аналитики многомерных временных рядов для краткосрочного прогнозирования эпидемиологического процесса. Для моделирования использовались реальные данные мониторинга эпидемии Грипп-ОРВИ в г. Киеве (официальный сайт МОЗ за ноябрь 2009 – январь 2010).

Ключевые слова: краткосрочное прогнозирование, эпидемиологические процессы, временные ряды, мониторинг, моделирование.

Abstract. A possibility of applying the methodology and analysis of multivariate time series for short-term forecasting of the epidemiological process is explored. Real data from monitoring the influenza epidemic SARS in Kiev (official website of Ministry of Health in November 2009 – January 2010) was used for the simulation.

Keywords: short-term forecasting, epidemiological processes, time series, monitoring, modeling.

1. Введение

Прогнозирование – необходимая процедура для принятия решения по предотвращению или выходу из критической ситуации (в данном случае – эпидемии). В данной статье рассматривается один из подходов к краткосрочному прогнозированию эпидемиологической обстановки, который, по мнению авторов, способен повысить эффективность системы прогнозирования и соответственно качество принимаемых решений.

2. Мониторинг заболеваемости гриппом

Эпидемиологическая ситуация, в соответствии с данными МОЗ [1], характеризуется совокупностью показателей, которые представляют собой заболеваемость групп населения по

№ з/д	Область	Кількість захворілих					З них госпіталізовано							
		Всього		дорослі		в т.ч. вагітні та діти до 18 років	Всього		знаходяться		дорослі		в т.ч. вагітні та діти до 18 років	
		з початку епідемії з 29.10.09	за останню добу	всього	в т.ч. вагітні та діти до 18 років		з початку епідемії з 29.10.09	за останню добу	в реанімації	на ШВЛ	всього	в т.ч. вагітні та діти до 18 років		
3	4	5	6	7	8	9	10	11	12	13	14			
1	АР Крим	108 345	415	187	2	228	4 605	62	14	6	22	2	40	
2	Вінницька	159 697	880	394	13	486	9 646	92	20	1	48	3	44	
3	Волинська	103 927	135	54	6	81	5 511	34	22	2	18	5	16	
4	Дніпропетровська	305 649	648	387	4	261	15 680	127	43	1	63	4	64	
5	Донецька	330 970	1 752	1 076	9	676	18 278	219	4	-	114	3	105	
6	Житомирська	100 740	367	194	3	173	7 663	84	2	-	26	1	58	
7	Закарпатська	82 893	209	101	3	108	6 230	28	-	-	10	-	18	
8	Запорізька	145 747	775	374	3	401	5 844	92	25	6	63	3	29	

Рис. 1. Фрагмент данных о заболеваемости гриппом, опубликованных на сайте МОЗ во время эпидемии 2009–2010 гг.

различным признакам (рис. 1):

- по возрасту (дети до 18 лет, взрослые);
- по группам риска (например, беременные);
- по тяжести заболевания (число госпитализируемых).

По этим данным определяется характер эпидемиологической ситуации – эпидемия, состояние, близкое к эпиде-

мии, отсутствие эпидемии. Данные фиксируются ежедневно.

Предполагаем рассматривать данные мониторинга заболеваемости в виде многомерного (векторного) временного ряда (рис. 2).

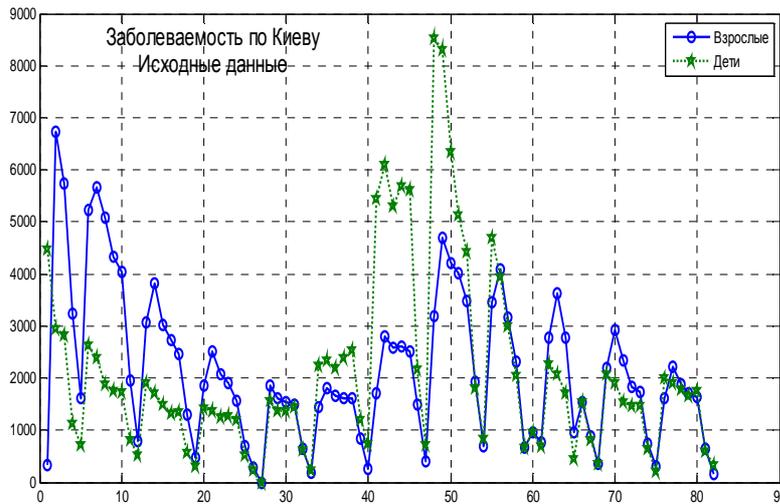


Рис. 2. Данные сайта МОЗ, представленные в виде временного ряда

заболеваемости как процесса. Период эпидемии исключается из рассмотрения. Поэтому этот подход не используется для решения задач прогнозирования. Многолетнее использование метода говорит об адекватности вероятностного подхода к моделированию показателей заболеваемости.

Динамика процесса заболеваемости учитывалась при создании моделей эпидемий. Подход, описанный у Бейли [3], в дальнейшем был обобщен для создания SIR-моделей [4, 5]. Суть подхода состоит в том, что определяется множество состояний индивидуума. Например, множество {чувствителен к инфекции, инфицирован (болен), выздоровел (нечувствителен к инфекции)} соответствует классической SIR-модели (*susceptible, infectious, recovered*). Описываются также правила и вероятности перехода из состояния в состояние. На основе данных предположений создается система дифференциальных уравнений, где искомые функции времени – количество индивидуумов, находящихся в каждом из определенных состояний. Пример применения модели SIR-класса – эпидемиологическая модель распространения туберкулеза [5].

Теоретически построенные SIR-модели имеют параметры, которые можно определить только на основании статистических данных. При моделировании эпидемии гриппа это вызывает сложности, обусловленные разнообразием и изменчивостью вирусов. Например, длительность эпидемии может меняться от 2 до 6 недель [2].

Наиболее объемное исследование с использованием статистических данных со всего СССР по эпидемиям гриппа было проведено Барояном и Рвачевым [7]. Отличие модели Барояна и Рвачева от класса SIR-моделей состоит в том, что состояние «болен» или «инфицирован» характеризуется функцией времени, прошедшего с момента инфицирования, которая и определяет способность индивида инфицировать «восприимчивых» или «чувствительных». В результате были получены нелинейные дифференциальные уравнения в частных производных. Применяя теоретическую модель на практике, столкнулись с тем, что параметры модели достоверно определяются только после момента, когда всплеск заболеваемости идет на спад. Поэтому эффект от применения моделирования авторы получили, прогнозируя эпидемиологический процесс в городах СССР по данным всплеска заболеваемости в городе, где эпидемия возникла раньше, и моделируя пассажирооборот между городами.

3. Известные модели эпидемиологического процесса заболеваемости гриппом

Вероятностные свойства показателей заболеваемости были использованы для разработки методологических указаний по вычислению эпидемических порогов [2]. Этот подход используется и в настоящее время в Украине [1]. В данной модели не используются динамические свойства показателя заболе-

Обзор методов моделирования эпидемиологических процессов и заболеваемости гриппом в том числе позволяет сделать следующие заключения:

- вероятностный подход к описанию свойств показателей заболеваемости не вызывает сомнения;
- динамические модели в виде нелинейных дифференциальных уравнений хорошо аппроксимируют всплески заболеваний гриппом во время эпидемий, однако параметры теоретических моделей определяются на основе статистических данных, и для гриппа это вызывает большие сложности, связанные с изменчивой природой вируса;
- модели описывают либо период между всплесками заболевания гриппом либо всплеск заболеваемости во время эпидемии;
- взаимозависимость между показателями заболеваемости не исследовалась и не использовалась для прогнозирования.

4. Анализ временных рядов, описывающих эпидемическую обстановку

Рассмотрим целесообразность применения методов анализа временных рядов для данных, характеризующих эпидемическую обстановку. Данный подход [8, 9] состоит в идентификации модели ряда, описывающей корреляционную зависимость между отсчетами, и позволяет учесть:

- стохастический характер исходных данных;
- одновременно эпидемиологический период и период между эпидемиями;
- динамику процессов изменения показателей заболеваемости во времени;
- взаимосвязь показателей и рассмотрение эпидемической обстановки как единого процесса.

Все реальные процессы в природе, как и эпидпроцесс, безусловно, являются нелинейными динамическими процессами. Однако на практике чаще используют линейное приближение из-за простоты и хороших результатов во многих ситуациях [8,9]. Прогнозируемое значение временного ряда в данном случае представляет собой линейную функцию текущего и предшествующих значений процесса, т.е. реализацию разностного уравнения вида

$$\hat{y}(t+T) = \sum_{i=0}^n \alpha_i \cdot y(t-i \cdot T), \quad (1)$$

где $\hat{y}(t+T)$ – прогнозируемое значение, $y(t)$ – исследуемый процесс, T – период дискретизации, α_i – коэффициенты, которые полностью определяются выбором модели динамического процесса. Если процесс является векторным, то α_i являются матрицами коэффициентов.

Для анализа временных рядов авторами был использован инструментальный Matlab [10], который позволяет исследовать три вида моделей: AR (авторегрессионную) модель, модель ARMA (авторегрессионную скользящего среднего) и линейную модель в пространстве состояний.

AR-модель имеет представление:

$$A(q)y(t) = e(t), \quad (2)$$

где $y(t)$ – исследуемый процесс, $e(t)$ – белый шум, q – оператор сдвига, $A(q)$ – полином. То есть соответствующее разностное уравнение имеет вид

$$a_0 \cdot y(t) + a_1 \cdot y(t-T) + \dots + a_n \cdot y(t-nT) = e(t), \quad (3)$$

где T – период дискретизации, n – порядок полинома A .

ARMA-модель представлена в виде

$$A(q)y(t) = B(q)e(t), \quad (4)$$

где $y(t)$ – исследуемый процесс, $e(t)$ – белый шум, q – оператор сдвига, $A(q)$ и $B(q)$ – полиномы.

Линейная модель в пространстве состояний описывается уравнениями:

$$x(t+T) = Ax(t) + Ke(t), \quad (5)$$

$$y(t) = Cx(t) + e(t),$$

где $x(t)$ – вектор состояний, $y(t)$ – исследуемый процесс, $e(t)$ – белый шум, A , K и C – числовые матрицы.

Для сравнения моделей в системе Matlab предлагается использовать следующий критерий степени сходства исходных (фактических) данных и прогнозируемых:

$$r^2 = \left(1 - \frac{\sum_{t=T}^{nT} (y(t) - y^s(t))^2}{\sum_{t=T}^{nT} (y(t) - \bar{y})^2}\right) \cdot 100\%, \quad (6)$$

где $y(t)$ – исследуемый процесс, T – период дискретизации, n – число отсчетов во временном интервале, для которого построено приближение. В некоторых источниках критерий называют коэффициентом детерминации.

5. Модельные исследования

Моделирование проводилось на данных мониторинга заболеваемости гриппом/ОРЗ, опубликованных на сайте МОЗ, с 11.2009 по 02.2010.

Для повышения информативности данных была проведена следующая предварительная обработка:

- компенсация грубых ошибок;
- определение коэффициентов, компенсирующих колебания заболеваемости в зависимости от дня недели;
- интерполяция данных, пропавших из-за праздников.

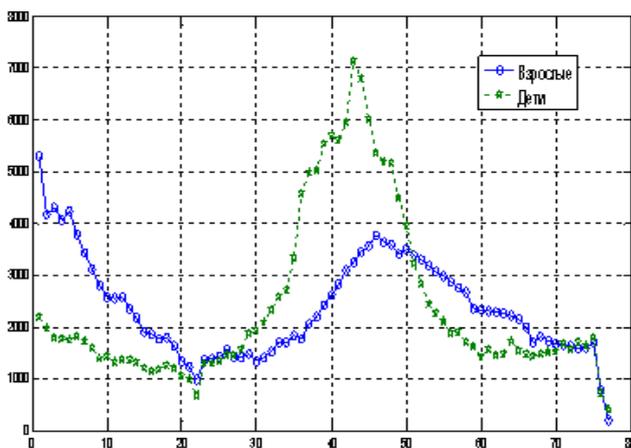


Рис. 3. Данные заболеваемости по Киеву после предварительной обработки

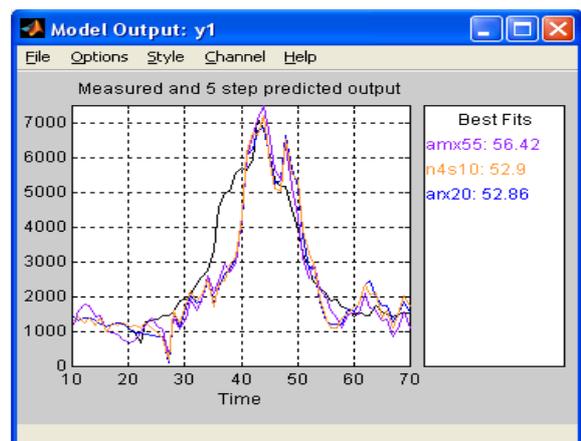


Рис. 4. Скалярная модель

Для сравнения результатов различных подходов к прогнозированию был проведен анализ как векторного, так и скалярного представления исследуемого процесса с исполь-

зованием моделей (2) – (5). Поскольку рассматриваемые методы разработаны для стационарных временных рядов, была проведена стандартная процедура исключения трендов.

На рис. 4. представлен результат прогнозирования, основанный на скалярных моделях. Показатель – заболевание детей до 18 лет. Сплошная линия – исходный временной ряд, пунктирная – прогноз на 5 дней. В правом окне на скриншоте – значения критерия (6) для каждого приближения.

В одномерном случае (рис. 4) качество прогноза по различным моделям приблизительно одинаковое. Порядок предпочтения – ARMA(4), линейная модель в пространстве состояний(5), AR(2). Разрядность моделей (степень полинома $A(q)$), соответствующая наилучшему приближению, равна 20.

Прогнозирование на основе многомерной модели значительно предпочтительнее, чем использование скалярной модели, что демонстрирует пример, приведенный на рис. 5. Векторный процесс в данном случае составлен из двух показателей – заболеваемости детей до 18 лет (y_1) и заболеваемости взрослых (y_2). Исходные данные о заболеваемости представлены сплошной линией взрослых (справа) и детей (слева), прогноз – пунктирная линия.

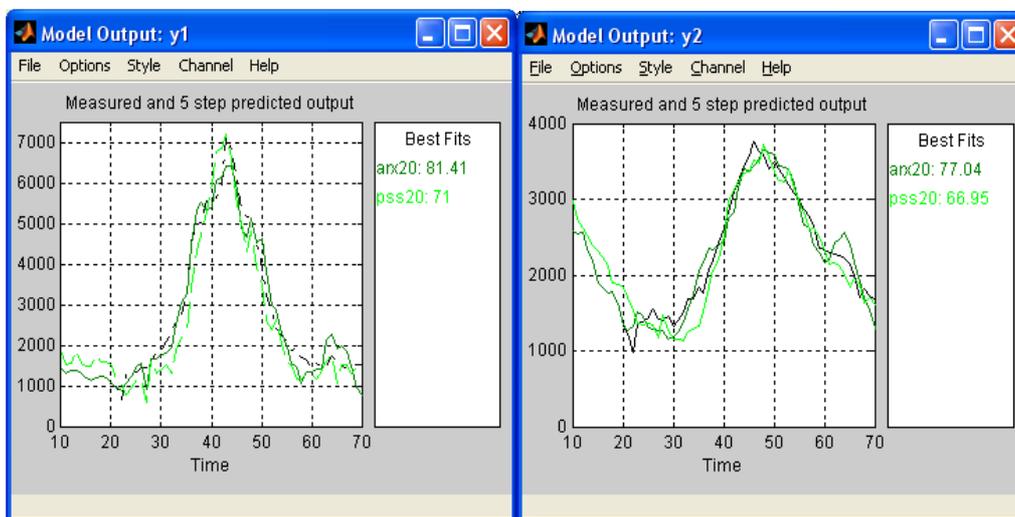


Рис. 5. Многомерный прогноз

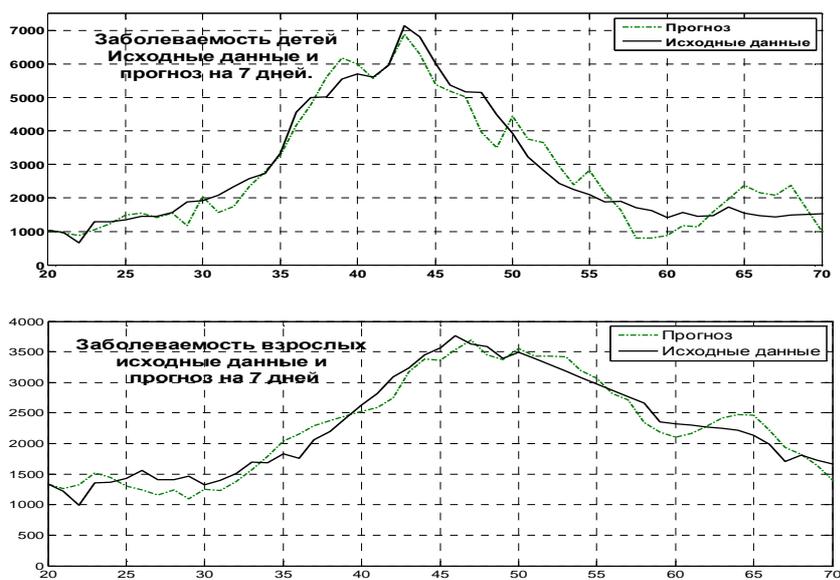


Рис. 6. Векторный прогноз на 7 дней

При двумерном прогнозе модель AR(2) показывает лучший результат, чем линейная модель в пространстве состояний (5). Разрядность моделей (степень полинома $A(q)$), соответствующая наилучшему приближению, как и в одномерном случае, равна 20.

Моделирование на различных данных позволяет сделать вывод, что наибольший интервал прогноза для модели AR

соответствует 7 дням (рис. 6).

6. Выводы

1. Анализ заболеваемости как временного ряда с помощью системы Matlab показал эффективность использования многомерных стохастических линейных динамических моделей для прогноза.
2. Подход может быть использован для краткосрочного прогноза ожидаемой эпидситуации по оперативным данным мониторинга.

СПИСОК ЛИТЕРАТУРЫ

1. Сайт МОЗ [Электронный ресурс]. – Режим доступа: <http://www.moz.gov.ua/ua/portal>.
2. Методические указания по оперативному анализу и прогнозированию эпидемической ситуации по гриппу и ОРЗ. – Л., 1999. – 59 с.
3. Бейли Н. Математика в медицине и биологии / Бейли Н. – М.: Мир, 1970. – 326 с.
4. Hethcote H. Qualitative analyses of communicable disease models / H. Hethcote // Math. Biosci. – 1976. – N 28. – P. 335 – 356.
5. Schaffer W.M. Parametric dependence in model epidemics. I: Contact-related parameters / W.M. Schaffer, T.V. Bronnikova // Journal of Biological Dynamics. – 2007. – Vol. 1, N 2. – P. 183 – 199.
6. Авилов К.К. Математические модели распространения и контроля туберкулеза [Электронный ресурс] / К.К. Авилов, А.А. Романюха // 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).
7. Бароян О.В. Математика и эпидемиология / О.В. Бароян, Л.А. Рвачев. – М.: Знание, 1977. – С. 63.
8. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – Т. 1. – 406 с.; М.: Мир, 1974. – Т. 2. – 194 с.
9. Льюинг Л. Идентификация систем. Теория пользователя / Льюинг Л. – М.: Наука, 1991. – 432 с.
10. MATLAB на сайте разработчика The MathWorks [Электронный ресурс]. – Режим доступа: <http://www.mathworks.com/products/matlab>.

Стаття надійшла до редакції 13.01.2011