

Технічні науки

УДК 519.8

**Світлова Олена Костянтинівна**

*студентка*

*Навчально-наукового комплексу «Інститут прикладного системного аналізу»*

*Національного технічного університету України*

*«Київський політехнічний інститут імені Ігоря Сікорського»*

**Светлова Елена Константиновна**

*студентка*

*Учебно-научного комплекса «Институт прикладного системного анализа»*

*Национального технического университета Украины*

*«Киевский политехнический институт имени Игоря Сикорского»*

**Svietlova Olena**

*Student of the*

*National Technical University of Ukraine*

*“Igor Sikorsky Kyiv Polytechnic Institute”*

**ПРОГНОЗУВАННЯ ПОШИРЕННЯ ІНФЕКЦІЙНИХ ЗАХВОРЮВАНЬ  
ПРОГНОЗИРОВАНИЕ РАСПРОСТРАНЕНИЯ ИНФЕКЦИОННЫХ  
ЗАБОЛЕВАНИЙ**

**PREDICTION THE SPREAD OF INFECTIOUS DISEASES**

*Анотація.* У статті розглянуті основні методи прогнозування поширення інфекційних захворювань — поточкову оцінку, регресійні моделі, байєсовські мережі та штучні нейронні мережі. Наведено особливості застосування, переваги та недоліки кожного з розглянутих методів.

*Ключові слова:* методи прогнозування, інфекційні захворювання, поточкова оцінка, регресійні моделі, байєсовські мережі.

*Аннотация.* В статье рассмотрены основные методы прогнозирования распространения инфекционных заболеваний - поточечную

оценку, регрессионные модели, байесовские сети и искусственные нейронные сети. Приведены особенности применения, преимущества и недостатки каждого из рассмотренных методов.

**Ключевые слова:** методы прогнозирования, инфекционные заболевания, поточечная оценка, регрессионные модели, байесовские сети.

**Summary.** The article considers the main methods for predicting the spread of infectious diseases - point evaluation, regression models, Bayesian networks and artificial neural networks. The features of the application, advantages and disadvantages of each of the methods considered are given.

**Key words:** prediction methods, infectious diseases, point estimation, regression models, Bayesian networks.

Методи прогнозування інфекційної захворюваності активно розвиваються з початку ХХ століття. В останні роки число робіт на цю тему стрімко зростає завдяки розвитку інформаційних систем нагляду і появи великих обсягів статистики, доступної для аналізу. Епідеміологічні прогнози виконуються для різних термінів і в залежності від них служать різним цілям. У цій статті представлені найбільш популярні методи.

## 1. Поточкова оцінка

Більшість інфекційних захворювань мають сезонний характер та повторюються з певною періодичністю.

Приклад 1.

Нехай  $T$  — період сезонності, тоді прогноз  $\hat{y}_t, \hat{y}_{t+1}, \dots, \hat{y}_{t-1+T}$  може бути обчислений для кожного перерізу процесу захворюваності на основі множин відомих значень  $\{y_{t-T}, y_{t-2T}, y_{t-3T}, \dots\}, \dots, \{y_{t-1}, y_{t-1-T}, y_{t-1-2T}, \dots\}$  відповідно.

Один з популярних способів отримання оцінок очікуваної захворюваності — просте усереднення:

$$\hat{y}_t = \sum_{j=1}^n y_{(t-jT)} / n, \text{ де } n \text{ — кількість доступних спостережень}$$

Цей підхід широко застосовується при обчисленні рівня ординарної захворюваності. Наприклад, діюча в Росії методика розрахунку епідемічних порогів грипу і ОРВІ передбачає усереднення даних за календарні тижні кожного року з однаковим номером протягом 5-10 останніх років. Такий спосіб прогнозування передбачає, що щорічна захворюваність носить постійний характер і спостереження протягом кожного нового року — наступна реалізація одного і того ж випадкового процесу.

У переважній більшості випадків це припущення не може бути дотримано. Якщо в динаміці захворюваності присутній виразний тренд, врахувати його можна за допомогою удосконалення підходу — поточкових лінійних оцінок (point-to-point linear estimates).

## **2. Регресійні моделі**

Класичний регресійний аналіз — напевно, найпопулярніший метод прогнозування захворюваності. Завдання регресії полягає в знаходженні оцінок невідомих параметрів і формуванні функціональної залежності між захворюваністю і факторами, її зумовлюють. Серед регресійних моделей слід виділити дві групи: неадаптивні моделі, для оцінки параметрів яких використовуються всі наявні дані, і адаптивні моделі, значення параметрів яких розраховуються на основі ковзаючого вікна спостережень.

Важливу роль при використанні адаптивних моделей грає ширина ковзаючого вікна — число останніх спостережень, на основі яких оцінюють параметри моделі. Її слід вибирати виходячи з терміну прогнозування. Як правило, адаптивні моделі застосовують при короткотерміновому прогнозуванні і ширина вікна не перевищує декількох місяців. Однак ширина вікна тісно пов'язана з видом регресійної залежності. Чим більше число невідомих параметрів, тим більше потрібно даних для отримання їх оцінок, і навпаки, чим ширше вікно, тим складніше може бути потрібно встановити для адекватного уявлення ділянки епідемічної кривої.

Наприклад, при прогнозуванні захворюваності на грип та ГРВІ на 12

тижні вперед за допомогою простої лінійної регресії ковзне вікно повинно уміщувати дані 4-6 тижнів. У загальному випадку, ширина вікна і навіть вид регресійної залежності можуть динамічно змінюватися протягом року або відповідно до характером епідемічної обстановки.

### **3. Байєсовські мережі**

Інший підхід до моделювання часових рядів, що забезпечує облік структури кореляції даних, — динамічні байєсовські мережі. Байєсовські мережі зображують у формі спрямованого графа, вершини якого відповідають змінним моделі, а ребра — ймовірнісним залежностям між ними, які задані певними законами розподілу. Після навчання, байєсовські мережі дозволяють оцінити ймовірність настання деякої події при спостерігаємій послідовності явищ. Байєсовські мережі сьогодні стрімко набирають популярність в різних областях знань і застосовуються до задачі прогнозування захворюваності переважно в простій формі прихованих марковських моделей (ПММ).

Основною ідеєю ПММ є зіставлення кожної випадкової величині  $Y_t$  з неспостерігаємою випадковою величиною  $S_t$ , що визначає умовний розподіл  $Y_t$ . ПММ припускають, що величина  $Y_t$  залежить тільки від значення прихованої змінної  $S_t$  в момент часу  $t$ . Послідовність  $S_t$  володіє марковською властивістю, тобто величина  $S_t$  залежить тільки від  $S_{t-1}$  (рис. 1а).

З цієї причини на поточний момент байєсовські мережі надають можливість тільки короткострокового прогнозування захворюваності. Більш того, найчастіше ПММ застосовуються лише для виявлення підвищеної захворюваності. При цьому будується проста модель з двома станами, що відповідають ординарній і епідемічній ситуації.

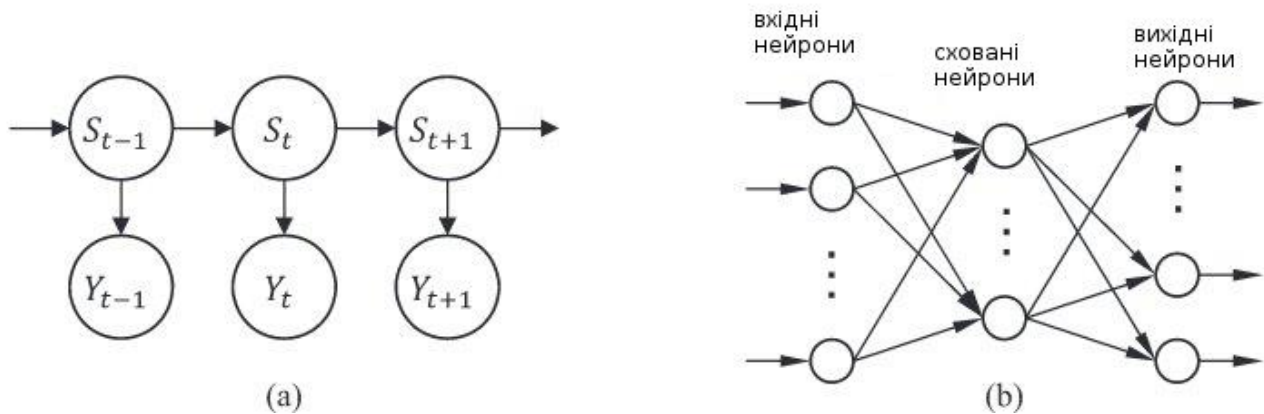


Рис. 1. Байєсові мережі

Розробка і аналіз байєсовських мереж і ПММ виконуються з використанням спеціалізованого програмного забезпечення.

#### 4. Штучні нейронні мережі

Важливий представник методів, що спираються на машинне навчання — штучні нейронні мережі (ШНМ). Методологія ШНМ широко відома і добре підходить для вирішення задач, аналітичне дослідження яких є важким. ШНМ представляє собою спрямований зважений граф, вершини якого моделюють функціонування біологічних нейронів (рис. 1b). Вершини приймають вхідні сигнали і при досить великому значенні їх зваженої суми перетворюють їх в вихідний сигнал. Навчання ШНМ полягає в обчисленні коефіцієнтів зв'язків між вершинами, що визначають силу вхідних сигналів, і виконується на основі емпіричних даних: статистики захворюваності і при наявності, значень факторів, що її зумовлюють.

Мета навчання ШНМ — визначення явних і неявних залежностей між вхідними і вихідними даними. Показники захворюваності, як правило, зашумлені, тому для навчання ШНМ потрібно порівняно велика їх кількість. Достатню для довгострокового прогнозування передісторію захворюваності ще тільки належить накопичити, тому поки ШНМ підходять скоріше для прогнозування на малі терміни.

В цілому ж ШНМ — вельми перспективний підхід. Методологія ШНМ реалізована в більшості систем комп'ютерної математики та статистичних

пакетів.

### **Література**

1. A practical guide for designing and conducting influenza disease burden studies. — World Health Organization, 2008. — 49 p. — URL: [http://www.wpro.who.int/emerging\\_diseases/documents/GuideforDesigningandConductingInfluenzaStudies/en/index.html](http://www.wpro.who.int/emerging_diseases/documents/GuideforDesigningandConductingInfluenzaStudies/en/index.html)
2. Киселев О.И., Маринич И.Г., Карпова Л.С., Ежлова Е.Б., Лазикова Г.Ф., Ватолина А.А. Методика расчета эпидемических порогов по гриппу и острым респираторным вирусным инфекциям по субъектам Российской Федерации. — М.: НИИ гриппа Северо-Западного отделения РАМН, 2010. — 88 с.
3. Serfling R.E. Methods for Current Statistical Analysis of Excess Pneumonia-influenza Deaths / Public Health Reports. — 1963. — Vol. 78, No 6. — P. 494–506.
4. Burkom H.S., Murphy S.P., Shmueli G. Automated Time Series Forecasting for Bio surveillance / Statistics in Medicine. — 2007. — Vol. 26, No 22. — P. 4202–4218.
5. Кизим Н.А., Доровской А.В. Анализ и прогнозирование тенденций заболеваемости населения Украины и Харьковской области / Проблемы экономики. — 2010. — No 3. — С. 39–44.
6. Watkins R.E., Eagleson S., Veenendaal B., Wright G., Plant A.J. Disease surveillance using a hidden Markov model / BMC Medical Informatics and Decision Making. — 2009. — Vol. 9, No 39.
7. Bai Y., Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy / Chaos, Solitons and Fractals. — 2005. — Vol. 26, No 2. — P. 559–569.