

УДК 519.246.8

DOI: 10.20998/2411-0558.2018.24.11

И. В. АНТОНОВА, канд. техн. наук, доц., НТУ "ХПИ",

Н. А. ЧИКИНА, канд. техн. наук, доц., НТУ "ХПИ"

СТРУКТУРНО-ЛИНГВИСТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ВРЕМЕННЫХ РЯДОВ

В работе рассматриваются вопросы, связанные с применением структурно-лингвистического подхода к анализу временных рядов, характеризующих динамику заболеваемости в Украине различными кожными патологиями. Нечеткое моделирование реализуется на основе анализа нечетких тенденций временного ряда. Такая модель позволяет получить качественную информацию о динамике показателей для использования ее в задачах прогнозирования. Ил.: 2. Библиогр.: 11 назв.

Ключевые слова: структурно-лингвистический подход; временной ряд; динамика заболеваемости; нечеткое моделирование; анализ нечетких тенденций, прогнозирование.

Постановка проблемы. Как правило, динамика социально-биологических систем характеризуется слабоструктурированным случайным процессом, а соответствующие им временные ряды (ВР) обладают высокой степенью неопределенности. Одной из причин этой неопределенности является слабовыраженная или отсутствующая тенденция в динамике развития. При статистическом подходе к моделированию таких ВР не удастся построить модель, которая адекватно описывала имеющийся ВР и давала бы приемлемую точность прогнозирования. Еще один недостаток статистических моделей состоит в том, что такие модели не предполагают лингвистической интерпретации как самих данных, так и прогнозных значений. Уровень требований, предъявляемых к результатам анализа ВР с высокой степенью неопределенности, способствует развитию нечеткого подхода к задаче прогнозирования. В случаях, когда прогнозное значение невозможно получить с требуемым уровнем точности, становится важной с практической точки зрения информация об ожидаемых тенденциях в динамике системы. Это делает актуальным анализ ВР с применением структурно-лингвистического подхода, реализующего анализ нечетких тенденций.

Анализ литературы. В представленной работе рассматриваются вопросы, связанные с анализом временных рядов, характеризующих динамику заболеваемости в Украине различными кожными патологиями. Источником информации о состоянии здоровья населения являются данные, содержащиеся в официальной статистической отчетности МОЗ

© И.В. Антонова, Н.А. Чикина, 2018

Украины [1]. В исследованиях [2] отмечалось отсутствие тренда в динамике показателей заболеваемости некоторыми кожными патологиями, что делает практически невозможным прогнозирование уровня заболеваемости статистическими методами.

Наличие указанных свойств в поведении ВР, описывающего динамику объекта изучения, определяет причины, по которым исследование рядов проводится методами нелинейной динамики (фрактальная геометрия, теория хаоса [3, 4] и др.). Наиболее известным приложением теории нелинейных систем с хаотическим поведением является прогнозирование динамики порождаемых ими ВР.

Обозначенная авторами проблематика тесно связана с задачами извлечения знаний из ВР на основе интеллектуальных технологий. Нечеткие модели ВР не претендуют на высокую точность прогнозирования. Они ориентированы на достаточно простое получение качественной информации о динамике изучаемой системы.

Нечеткое моделирование поведения сложных систем по имеющемуся ВР базируется на модели нечеткого динамического процесса, получившей название нечеткий ВР (НВР). Математическую основу НВР составляет, в частности, знаменитая теорема FAT (Fuzzy Approximation Theorem), доказанная Kosko B., в соответствии с которой любая функциональная зависимость, заданная на компактном множестве, может быть аппроксимирована нечеткой моделью [5]. Практически одновременно в [6] Wong L. показал, что нечеткие модели вида "Если-То" являются универсальным аппроксиматором, т.е. модель может приближать любую непрерывную функцию с произвольной точностью. Однако, доказательства аппроксимирующих свойств нечетких систем являются чисто теоретическими, поскольку не дают конструктивного метода построения нечеткой модели с требуемой точностью. Лингвистическим нечетким правилам соответствуют аппроксимирующие нечеткие модели Мамдани, точность в которых может быть повышена при условиях, указанных в [7].

Song Q. и Chissom B., Sah M. и Degtiarev K.Y. в работах [8, 9] предложили прогнозные модели НВР первого порядка, нечеткие значения которого представляют фазсифицированные первые разности исходного ВР. Таким образом был введен в изучение новый объект НВР – нечеткая тенденция, выражающая изменения в динамике системы в нечетких значениях ВР.

Целью настоящих исследований является изучение вопросов, связанных с построением модели временного ряда на основе анализа нечетких тенденций, позволяющей получить качественную информацию

о динамике показателя заболеваемости некоторым классом кожных патологий в Украине для использования ее в задачах прогнозирования.

Основной раздел. Под моделью ВР будем понимать формализованную процедуру, позволяющую по прошлым значениям $x(t_i)$, $i = \overline{1, n}$ ВР прогнозировать его будущие значения $Y(t)$, $t > t_n$.

Метод представления ВР в виде НВР основан на предположении о реальной возможности построения лингвистической интерпретации значений $x(t_i)$ $i = \overline{1, n}$ ВР, соответствующей понятию нечеткого множества. При этом НВР формируется в результате интервального качественного оценивания уровней ВР. Изменения в НВР рассматриваются как результат значимого или незначимого влияния факторов, в общем случае исследователю неизвестных. Значимые влияния факторов приводят к качественному изменению (нечетким приращениям) поведения ВР, которые получили название нечеткие тенденции.

Моделирование таких нечетких объектов, как значения исходного ВР и нечетких тенденций, основано на следующих положениях [10]:

- 1) модель НВР представлена распределением нечеткости с помощью лингвистической переменной с заданным количеством нечетких термов;
- 2) результат моделирования ВР: НВР значений (а) и НВР нечетких тенденций (б);
- 3) каждому значению ВР (а) и (б) соответствует определенный нечеткий терм.

Нечеткая тенденция содержит информацию о ВР, выраженную в значениях лингвистической переменной, таких как "Рост", "Снижение", "Стабильность" и др. Последовательность нечетких тенденций НВР порождает НВР с нечеткой тенденцией.

Пусть задан скалярный эквидистантный временной ряд $\{x(t_i)\}_{i=1}^n$, $x(t_i) \in X \subset R$, $i = \overline{1, n}$, измерения которого $x(t_i)$ в моменты наблюдений t_i , $i = \overline{1, n}$, характеризуют заболеваемость в Украине некоторым классом кожных патологий [1].

Главной частью модели прогноза НВР является база знаний, получаемая на основе анализа и обработки имеющихся данных в ВР $X(t) = \{x(t_i)\}_{i=1}^n$.

Для нечеткой входной переменной $X(t)$ – уровня заболеваемости – определим терм-множество со значениями: низкий (Н), средний (С),

высокий (В). Для прогнозируемого (выходного) параметра $Y(t)$, который также является нечеткой переменной, определим пятиуровневую градацию с таким терм-множеством: значительное уменьшение (ЗУ), уменьшение (У), стабильность (С), рост (Р), значительный рост (ЗР). Такой уровень градации значений выходного параметра обеспечивает достаточный уровень дифференциации прогноза.

Нечеткие шкалы терм-значений $X(t)$ изучаемого показателя динамики заболеваемости и терм-значений прогнозируемого показателя $Y(t)$ строятся на основе экспертных оценок.

Для иллюстрации метода обнаружения нечетких тенденций выберем ВР, характеризующий уровень заболеваемости в Украине различными кожными патологиями, изображенный на рис. 1.

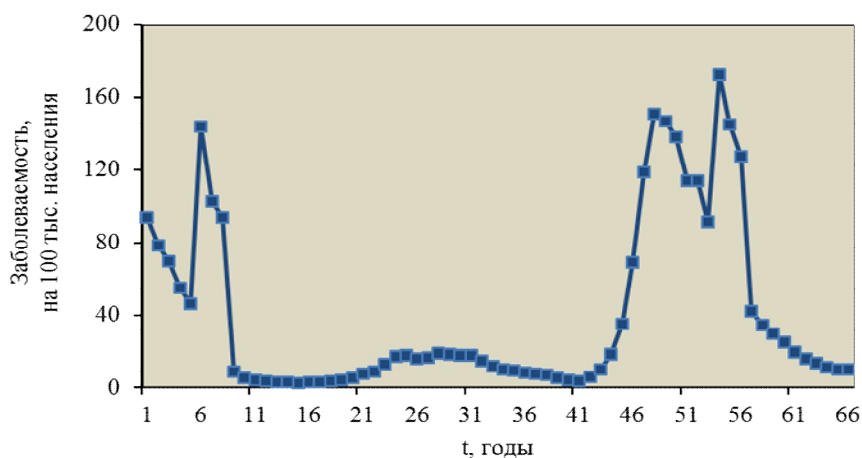


Рис. 1. Динамика показателя заболеваемости кожными патологиями в Украине

Алгоритм обнаружения нечетких тенденций НВР, используемый авторами, состоит в следующем [11]. Интервал наблюдения ВР с помощью нечетких ассоциативных правил разбивается на m интервалов ω_k , $k = \overline{1, m}$ монотонности $[t_{i_k}, t_{i_{k+1}}]$, при этом $t_{i_1} = t_1$, а $t_{i_{m+1}} = t_n$, т.е. интервалов его монотонного убывания или возрастания. Обозначим через $\Delta_1, \Delta_2, \dots, \Delta_m$ длины интервалов монотонности, т.е. $\Delta_k = t_{i_{k+1}} - t_{i_k}$, $k = \overline{1, m}$.

На каждом из этих интервалов $\omega_k (k = \overline{1, m})$ отрезок ВР аппроксимируется прямой $T_k (k = \overline{1, m})$.

Тангенс угла наклона прямой T_k к положительному направлению оси OX характеризует величину тренда ВР на участке ω_k , длина Δ_k участка ω_k монотонности ВР – его продолжительность. Один из вариантов такого разбиения представлен на рис. 2.

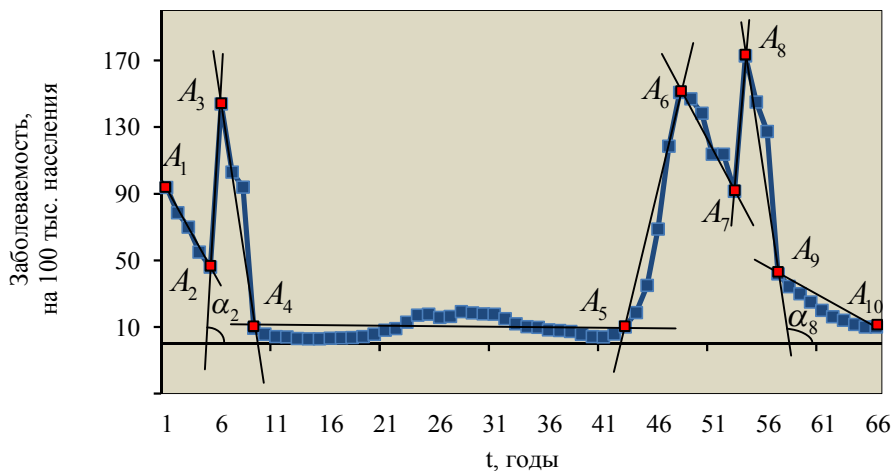


Рис. 2. Разбиение ВР на интервалы монотонности и соответствующие им аппроксимации прямыми линиями

Для формализации введенных признаков определим нечеткие переменные – "Тренд" и "Продолжительность". Нечеткой переменной "Тренд" соответствуют нечеткие термы $\alpha_1, \alpha_2, \dots, \alpha_m$ – тангенсы углов наклона прямых T_k , $k = \overline{1, m}$, а нечеткой переменной "Продолжительность" – нечеткие термы $\Delta_1, \Delta_2, \dots, \Delta_m$.

Единицей представления информации о ВР на каждом промежутке монотонности $[t_{i_k}, t_{i_{k+1}}]$, $k = \overline{1, m}$ является нечеткое высказывание вида:

$$I_k(t_{i_k}, t_{i_{k+1}}) = (Tr(t_{i_k}, t_{i_{k+1}}) = \alpha_k) \cap (t_{i_{k+1}} - t_{i_k} = \Delta_k). \quad (1)$$

Смысл этого высказывания состоит в следующем: на временном интервале $[t_{i_k}, t_{i_{k+1}}]$ с нечеткой продолжительностью Δ_k наблюдается тренд ВР с нечетким значением α_k . Это высказывание имеет также и геометрическую интерпретацию: информация о ВР представляет собой

нечеткое множество векторов $\overrightarrow{A_{i_k} A_{i_{k+1}}}$, характеризующих тренд временного ряда на интервалах $[t_{i_k}, t_{i_{k+1}}]$, $k = \overline{1, m}$.

Прогнозирование НБР осуществляется на основе алгоритма нечеткого логического вывода. База знаний формируется по промежуткам ω_k , $k = \overline{1, m}$ интервала наблюдения $[t_1, t_n]$. Единицей представления знаний в модели являются нечеткие правила вида:

$$\text{"Если } I_1 \text{ и } I_2 \text{ и } \dots \text{ и } I_m, \text{ то } Y(t_p) = I_{m+1}(t_n, t_p)\text{"}, \quad (2)$$

где I_k , $k = \overline{1, m+1}$ – нечеткое высказывание вида (1), t_p – точка прогноза.

Следует отметить, что наибольшими проблемами исследователя в реализации алгоритма нечетких тенденций являются выбор интервала наблюдения ВР, в наибольшей степени характеризующего динамику изучаемой системы, и выбор точек изменения нечетких тенденций.

Выводы. Идея использования аппарата нечеткой логики является перспективной для оценки информации, содержащейся во ВР. Метод прогнозирования нечетких тенденций ВР динамики показателей заболеваемости различными кожными патологиями в Украине обеспечивает получение более точных результатов прогноза и может быть рекомендован для проведения краткосрочных прогнозов.

Моделирование ВР с нечеткими тенденциями позволит прогнозировать не только числовые значения ВР, но и направления их изменения, а также получить информацию о поведении ВР в лингвистической форме.

Такой подход позволит обнаружить новые закономерности изучаемого случайного процесса и извлечь из ВР информацию, недоступную другим подходам к моделированию.

Список литературы:

1. Показники лікувально-профілактичної допомоги хворим шкірними і венеричними захворюваннями в Україні у 2015 році / відповідальний за випуск М. В. Голубчиков; ДЗ "Центр медичної статистики МОЗ України". – К., 2016. – 113 с.
2. Волкославская В.Н. О квазипериодичности в динамике заболеваемости сифилисом и основных направлениях организации борьбы с ИППП в Украине / В.Н. Волкославская, А.Л. Гутнев, Н.А. Чикина // Труды науч.-практ. конф. "Актуальные проблемы дерматологии и венерологии". – Одесса: ОГМУ, 2003. – С. 25-26.
3. Антонова И.В. Применение методов фрактального анализа к исследованию временных рядов / И.В. Антонова, Н.А. Чикина // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ", 2015. – № 32 (1141). – С. 4-10.
4. Антонова И.В. Применение характеристик квазициклов фазовых портретов в предпрогножном анализе временных рядов / И.В. Антонова, Н.А. Чикина // Вестник

- НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ", 2017. – № 21 (1243). – С. 5-13.
5. Kosko B. Fuzzy Systems as Universal Approximators / B. Kosko // IEEE Trans. on Computers, 1994. – Vol. 43. – № 11. – P. 1329-1333.
6. Wang L.X. Fuzzy systems are universal approximators / L.X. Wang // Proc. of the IEEE Int. Conf. on Fuzzy Systems. San Diego, 1992. – P. 1163-1169.
7. Кукса П.П. Обеспечение точности в нечетких системах / П.П. Кукса // Информатика и системы управления в XXI веке. – М.: МГТУ им. Н.Э. Баумана, 2004. – С. 152-156.
8. Song, Q. Fuzzy time series and its models / Q. Song, B. Chissom // Fuzzy sets and Systems, 1993. – № 54. – P. 269-273.
9. Sah M. Forecasting Enrollment Model Based on First Order / M. Sah, K.Y. Degtiarev // Proct. Int. Conf. Computational Intelligence (ICCI), 2004. – P. 375-378.
10. Афанасьева Т.В. Моделирование нечетких тенденций временных рядов / Т.В. Афанасьева. – Ульяновск: УлГТУ, 2013. – 215 с.
11. Тукаева З.М. Модель прогнозирования нечетких данных для решения бизнес задач предприятий ТЭК / З.М. Тукаева, И.З. Мухаметзянов // Управление экономическими системами: электронный научный журнал, 2013. – № 56, available at: www.uecs.ru/uecs56-562013/item/2299-2013-08-27-07-57-10 (accessed data 02.05.2018).

References:

1. Golubchikov, M.V. (2016), "*Indexes of medical and preventive help to the patients with skin and venereal diseases at Ukraine in 2015*", state establishment "Center of medical statistics of ministry of health Ukraine", Kyiv, 113 p.
2. Volkoslavskaya, V.N., Gutnev, A.L., and Chikina, N.A. (2003), "About quasi periodicity in the dynamics of the syphilis morbidity and basic directions of organization of fight with IPPP in Ukraine", *Actual problems of dermatology and venereology*, Proceedings of the Science and Practical Conference, Odessa, pp. 25-26.
3. Chikina, N.A., and Antonova, I.V. (2015), "Application of fractal analysis methods to the study of time series", *Herald of NTU "KhPI", Series: Informatics and modeling*, Kharkov, NTU "KhPI", No 32 (1141), pp. 4-10.
4. Chikina, N.A., and Antonova, I.V. (2017), "Application of quasi-cycle characteristics of phase portraits in the prediction of time series analysis", *Herald of NTU "KhPI", Series: Informatics and modeling*, Kharkov, NTU "KhPI", No 21 (1243), pp. 5-13.
5. Kosko, B. (1994), "Fuzzy Systems as Universal Approximators", *IEEE Trans. on Computers*, Vol. 43, No. 11, pp. 1329-1333.
6. Wang L. X. (1992), "Fuzzy systems are universal approximators", *Proc. of the IEEE Int. Conf. on Fuzzy Systems*, San Diego, pp. 1163-1169.
7. Kouxa, P.P. (2004), "Approximation Accuracy of Fuzzy Systems", *Informatics and Control Systems in 21st century*, Moscow, BMSTU Press, pp. 152-156.
8. Song, Q., and Chissom, B. (1993), "Fuzzy time series and its models", *Fuzzy sets and Systems*, No 54, pp. 269-273.
9. Sah, M., and Degtiarev, K.Y. (2004), "Forecasting Enrollment Model Based on First Order", *Proct. Int. Conf. Computational Intelligence (ICCI)*, pp. 375-378.
10. Afanas'yeva, T.V. (2013), *Simulation of fuzzy tendencies of time series*, Ulyanovsk, 215 p.
11. Tukaeva, Z.M. and Mukhametzyanov, I.Z. (2013), "Model of forecasting of fuzzy data for a decision of business tasks of TEK enterprises", *Management of the economic systems: electronic scientific magazine*, available at: www.uecs.ru/uecs56-562013/item/2299-2013-08-27-07-57-10 (accessed data 02.05.2018).

*Статью представил д-р техн. наук, проф. НТУ "ХПИ"
Поворознюк А.И.*

Поступила (received) 07.05.2018

Повторно 14.05.2018

Antonova Irina, PhD Tech., Associate Professor
National Technical University "Kharkiv Polytechnic Institute"
Str. Kirpicheva, 21, Kharkov, Ukraine, 61002
Tel.: (057) 707-60-87, e-mail: antonova2601@gmail.com
ORCID ID: 0000-0002-1268-8223

Chikina Natalia, PhD Tech., Associate Professor
National Technical University "Kharkiv Polytechnic Institute"
Str. Kirpicheva, 21, Kharkov, Ukraine, 61002
Tel.: (057) 707-66-93, e-mail: nachikina56@gmail.com
ORCID ID: 0000-0002-1746-1271

УДК 519.246.8

Структурно-лінгвістичний підхід до аналізу часових рядів / Антонова І.В., Чікіна Н.О. // Вісник НТУ "ХПІ". Серія: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2018. – № 24 (1300). – С. 130 – 138.

У роботі розглядаються питання, пов'язані із застосуванням структурно-лінгвістичного підходу до аналізу часових рядів, що характеризують динаміку захворюваності в Україні різними шкірними патологіями. Нечітке моделювання реалізується на основі аналізу нечітких тенденцій часового ряду. Така модель дозволяє отримати якісну інформацію про динаміку показників для використання її в завданнях прогнозування. Іл.: 2. Бібліогр.: 11 назв.

Ключові слова: структурно-лінгвістичний підхід; часовий ряд; динаміка захворюваності; нечітке моделювання; аналіз нечітких тенденцій; прогнозування.

УДК 519.246.8

Структурно-лингвистический подход к анализу временных рядов / Антонова И.В., Чикина Н.А. // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2018. – № 24 (1300). – С. 130 – 138.

В работе рассматриваются вопросы, связанные с применением структурно-лингвистического подхода к анализу временных рядов, характеризующих динамику заболеваемости в Украине различными кожными патологиями. Нечеткое моделирование реализуется на основе анализа нечетких тенденций временного ряда. Такая модель позволяет получить качественную информацию о динамике показателей для использования ее в задачах прогнозирования. Ил.: 2. Библиогр.: 11 назв.

Ключевые слова: структурно-лингвистический подход; временной ряд; динамика заболеваемости; нечеткое моделирование; анализ нечетких тенденций; прогнозирование.

UDC 519.246.8

Structurally linguistic method of the analysis of time series / Antonova I.V., Chikina N.A. // Herald of NTU "KhPI". Series: Informatics and modeling. – Kharkov: NTU "KhPI". – 2018. – No. 24 (1300). – P. 130 – 138.

The problems, linked with the use of the structurally linguistic method of the analysis of time series, which characterize the morbidity dynamics in Ukraine by different skin pathologies, are considered. Fuzzy simulation is realized on the basis of fuzzy tendencies analysis of time series. Such model allows getting quality information about the dynamics of the indexes to use it in the tasks of forecasting. Figs.: 2. Refs.: 11 titles.

Keywords: structurally linguistic method; time series; dynamics of morbidity; fuzzy simulation; analysis of fuzzy tendencies; forecasting.