

УДК 519.246.8

Я.В. Пирогова, Т.Г. Ємел'яненко

Дніпропетровський національний університет ім. Олеся Гончара

РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ КОРОТКОСТРОКОВОГО ПРОГНОЗУВАННЯ НЕЧІТКИХ ЧАСОВИХ РЯДІВ

Створено інформаційну технологію для побудови прогнозів нечітких часових рядів на базі нечіткої логіки з використанням генетичного алгоритму. Розглянутий підхід покладено до основи розробленого програмного продукту *Fuzzy_Forecassting*, який створено у середовищі Eclipse на мові програмування Java.

Ключові слова: нечітка логіка, нечіткий часовий ряд, прогнозування, генетичний алгоритм, програмний продукт, гідрогеохімічний моніторинг.

Создана информационная технология для прогнозирования нечетких временных рядов на базе нечеткой логики с использованием генетического алгоритма. Рассмотренный подход лежит в основе разработанного в среде Eclipse программного продукта *Fuzzy_Forecassting* на языке программирования Java.

Ключевые слова: нечеткая логика, нечеткий временной ряд, прогнозирование, генетический алгоритм, программный продукт, гидрогеохимический мониторинг.

It was created information technology to make forecast for fuzzy time series based on fuzzy logic and genetic algorithm. This approach was used in the developed software product *Fuzzy_Forecassting* which was created in Eclipse on the programming language Java.

Key word: fuzzy logics, fuzzy time series, forecasting, genetic algorithm, software product, hydrogeochemical monitoring.

Вступ. У галузі розробки інформаційних технологій актуальною є задача прогнозування процесу, що аналізується. Це обумовлено тим, що ефективно управління під час рішення як технічних, так і економічних задач, неможливе без достовірного й оперативного прогнозування. У теорії і практиці за минулі роки накопичено значний набір різноманітних методів розробки прогнозів. За оцінками вчених, нараховується понад 150 різних методів прогнозування; на практиці ж в якості основних використовуються лише 15–20. Розвиток

інформатики і засобів обчислювальної техніки створює можливість розширення кола методів прогнозування та їх вдосконалення [1].

Мабуть, найбільш вражаючою властивістю людського інтелекту є здатність приймати правильні рішення в обстановці неповної і нечіткої інформації. Побудова моделей наближених міркувань людини і використання їх у комп'ютерних системах представляє сьогодні одне з дуже цікавих наукових питань. Значне просування в цьому напрямку зроблено професором Каліфорнійського університету (Берклі) Лотфі А. Заде (Lotfi A. Zadeh). Його роботи заклали основи моделювання інтелектуальної діяльності людини і стали початковим поштовхом до розвитку нової математичної теорії. Заде розширив класичне канторовське поняття множини, припустивши, що характеристична функція (функція приналежності елемента множині) може приймати будь-які значення в інтервалі $[0; 1]$, а не тільки значення 0 або 1. Такі множини були названі їм нечіткими (fuzzy).

Актуальною є задача дослідження можливості застосування нечіткого підходу до прогнозування техногенного навантаження на територіях гірничо-збагачувальних комбінатів (ГЗК).

Аналіз останніх досліджень і публікацій. Прогнозування, яке базується на нечітких часових рядах (НЧР), притягує увагу дослідників протягом останніх 15 років. Більшість опублікованих до теперішнього часу праць використовували в якості тестової послідовності даних дані реєстрації студентів в Університеті шт. Алабама за період у майже 20 років. Дослідники Сонг та Чісс (Song-Chissom) [6 – 8] розглядали стаціонарні (time-invariant) та змінні у часі (time-variant) НЧР моделі прогнозу, які забезпечили помилки на рівні 3.18 % та 4.37 % (вікно прогнозування дорівнює 4), відповідно. Стаціонарна НЧР модель, яка була запропонована Ченом (Chen) [9] та протестована на тих самих даних, практично не змінила похибку прогнозу Сонга та Чісса, (3.23 %), а модифікований підхід Шаха і Дегтярьова (Şah, Degtiarev) [10] дозволив трохи покращити результати, знизивши середню помилку до 2.42 %. Одночасно, змінна у часі НЧР модель Хванга, Чена та Лі (Hwang-Chen-Lee) [11], залишаючись простою з точки зору матричних обчислень, забезпечила відносну похибку близько 3.12 %.

Найчастіше нечіткі часові ряди використовуються під час прогнозування даних фінансових ринків. Значно менша кількість робіт присвячена прогнозуванню екологічних даних з використанням нечіткого підходу [11 – 15]. Застосуванню нечітких часових рядів для розв'язання задачі прогнозування даних гідрогеохімічного моніторингу не приділено достатньої уваги.

Постановка задачі. Задано часовий ряд , що містить інформацію про концентрацію хімічної речовини в пробах ґрунтових вод на техногенно навантаженій території Північного гірничо-збагачувального комбінату (ГЗК). Необхідно створити інформаційну технологію для побудови прогнозу, тобто продовження часового ряду у вигляді

$$u_t, t = \overline{1, N+1}.$$

Основний матеріал. Останнім часом з'явилося багато різноманітних методів прогнозування, які основані на теорії нечітких множин. Вони відрізняються тим, що використовують різні алгоритми нечіткого виводу. Але ці методи мають певні обмеження. Найвідоміші алгоритми: Мамдані, Цукамото, Ларсена, Сугено. У цілому, зміст цих алгоритмів зводиться до наступних етапів:

1. Фазифікація – це процедура знаходження значень функції належності вхідних лінгвістичних змінних на основі звичайних (не нечітких) даних. Іншими словами, треба вибрати вид функції належності та визначити проміжки, на яких вхідна лінгвістична змінна асоціюється із заданими на даному етапі лінгвістичними оцінками.

2. Побудова нечіткої бази знань – запис набору нечітких правил (експертних висловлювань), що мають вид «ЯКЩО – ТОДІ», що описують зміну рівнів часового ряду та факторів, які впливають на показники, що прогножуються. База знань систем нечіткого виводу призначена для формального представлення емпіричних знань та знань експертів у той чи іншій проблемній галузі.

3. Агрегування – це процедура визначення степеня істинності умов за кожним із правил системи нечіткого виводу.

4. Активізація у системах нечіткого виводу – це процедура знаходження степені істинності кожного із підзаключень правил нечітких продукцій.

5. Аккумуляція – процедура або процес знаходження функції належності для кожної із вихідних лінгвістичних змінних.

6. Дефазифікація – це процедура або процес знаходження звичайного (не нечіткого) значення для кожної вихідної лінгвістичної змінної множини. Для дефазифікації можуть бути використані такі методи: центра тяжіння, центра тяжіння для одноточкових множин, центра площини, лівого та правого модальних значень, центр максимумів [2].

Для збільшення якості прогнозів необхідно використовувати механізм налаштування моделі.

Схема реалізації цієї методики виглядає так.

Алгоритм 1

Крок 1. Визначення універсальної множини U . Нехай D_{\min} та D_{\max} найменше та найбільше значення вихідних даних. З метою отримання більш зручних меж інтервалу треба вибрати додатні числа D_1 і D_2 . Тоді універсальну множину U можливо представити у вигляді $U : U = [D_{\min} - D_1; D_{\max} + D_2]$.

Крок 2. Розіб'ємо універсальну множину на k рівних інтервалів.

Крок 3. Визначимо середні точки інтервалів u_{icp} .

Крок 4. Визначимо нечіткі множини універсальної множини U . Визначимо A_i ($i = 1, \dots, k$) лінгвістичні значення лінгвістичної змінної за формулою

$$\mu_{A_i}(u_i) = \frac{1}{1 + (c \cdot (v - u_{icp}))^2}, \quad (1)$$

де v – варіація, c – стала величина, яка вибирається так, щоб забезпечити належність $\mu_{A_i}(u_i)$ проміжку $[0; 1]$.

Крок 5. Фазифікація варіацій, обчислених на кроці 1. Якщо для кварталу i варіація буде v_i , v_i належить u_j , то для u_j (u_j належить U – інтервали універсальної множини U) функція приналежності $\mu_{A_i}(u_i)$ обчислюється за формулою (1) з урахуванням $v = v_i$, тобто з універсальної множини розглядається той інтервал, до якого потрапляє варіація, що розглядається.

Крок 6. Треба вибрати базис w ($1 \leq w \leq l$, де l – кількість попередніх кварталів, які включені у прогноз). З урахуванням базису, обчислюється матриця нечітких відносин $R^w(t)$, на основі якої робиться прогноз. З цією метою після вибору w будується операційна матриця $i \times j$ $O^w(t)$ (i – число рядків, що відповідають кількості кварталів у послідовності $t-2, t-3, \dots, t-w$; j – число стовпчиків, що відповідає кількості інтервалів варіацій) та матриця-критерій $1 \times j$ $K(t)$ для кварталу, що прогнозується, t (матриця-рядок, що відповідає нечіткій варіації за квартал $t-1$).

Крок 7. Для дефазифікації результатів, отриманих на 6-му кроці, пропонується така формула

$$v(t) = \frac{\sum \mu_t(u_i) \cdot u_{icp}}{\sum \mu_t(u_i)}$$

де $\mu_t(u_i)$ – обчислені значення функції приналежності для кварталу, для якого проводиться прогноз.

Отже, в результаті отримано значення прогнозу варіації величини на наступний квартал і для того щоб отримати значення прогнозу самої величини, необхідно до поточного значення величини додати значення спрогнозованої варіації [3].

Нечіткі методи дуже успішно застосовуються, отримані результати мають високий ступень інтерпретації. Але точність автоматично отриманих правил дуже низька, часто виникають ситуації, коли необхідно використовувати допомогу експертів. Для виправлення цього недоліку, можливо використовувати еволюційні методи, щоб будувати та налаштовувати набір нечітких правил. Питаннями побудови та налаштуванням нечітких правил займається напрямок генетичних нечітких систем (genetic fuzzy system). Дослідження в цьому напрямку є дуже важливою актуальною науковою задачею, під час розв'язку якої можна отримати важливі результати як у теоретичному так і практичному плані [4].

Розглянемо один з сучасних методів прогнозування НЧР з використанням генетичного алгоритму. Цей алгоритм може бути представлений наступним чином.

Крок 1. Визначення універсальної множини U . Нехай D_{\min} та D_{\max} найменше та найбільше значення вихідних даних. З метою отримання більш зручних меж інтервалу треба вибрати додатні числа D_1 і D_2 . Тоді універсальну множину U можна представити у вигляді $U : U = [D_{\min} - D_1; D_{\max} + D_2]$.

Крок 2. розіб'ємо універсальну множину на 7 інтервалів: $[D_{\min} - D_1; x_1], [x_1; x_2], [x_2; x_3], [x_3; x_4], [x_4; x_5], [x_5; x_6], [x_6; D_{\max} + D_2]$, де $x_1, x_2, x_3, x_4, x_5, x_6$ – цілі числа та $x_1 < x_2 < x_3 < x_4 < x_5 < x_6$.

Крок 3. Визначимо хромосому. Нехай кожна хромосома складається з 6 генів, як показано нижче

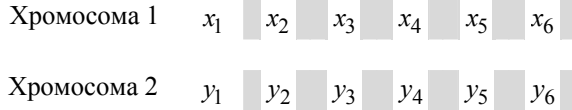
$$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6,$$

де $x_1, x_2, x_3, x_4, x_5, x_6$ – цілі числа.

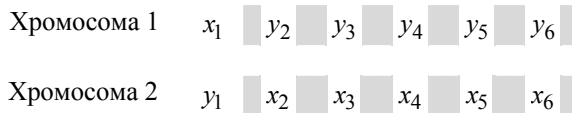
Під час ініціалізації випадковим чином генерується кожний ген. Випадковим чином згенеруємо популяцію N хромосом.

Крок 4. Значення прогнозу може бути знайдено використовуючи крок 3 – крок 6 з Алгоритму 1.

Крок 5. Кроссовер. Випадковим чином виберемо дві хромосоми з популяції, щоб провести процедуру кроссоверу. Під час процедури кроссоверу система випадково вибирає точку кроссоверу (ціле число від 1 до 5).



Після кроссоверу:



Крок 6. Мутація. Випадковим чином вибираємо хромосому з популяції та випадковим чином вибираємо номер гена у хромосомі. Після проведення операцій мутації та кроссоверу, для кожної хромосоми проводимо процедуру прогнозування. Коли отримано всі значення прогнозів, можна обчислити помилку прогнозу. Значення помилки буде використовуватися як значення фітнесу для відповідної хромосоми. Зрозуміло, чим менше помилка, тим краще пристосована хромосома.

Крок 7. Вибираємо M ($M < N$) хромосом із популяції з найменшими значеннями помилки, і ще $(N - M)$ хромосом генеруємо випадковим чином. Аналогічним чином для нової популяції повторюємо кроки 5 – 7. Таким чином, треба породити P популяцій. Хромосома, яка буде мати найменше значення помилки і буде найкращим розв'язком задачі прогнозування НЧР [5].

Розглянуті алгоритми покладені до основи розробленого програмного продукту Fuzzy_Forecasting. Середовище розробки програмного продукту – Eclipse. Мова програмування – Java. Мінімальні вимоги до системи: операційна система Windows XP/7; Linux. Програмний продукт спроектовано за шаблоном Модель-Вигляд-Контролер (Model – View – Controller). На рис. 1 наведено UML діаграму класів, а на рис. 2 – діаграму варіантів використання.

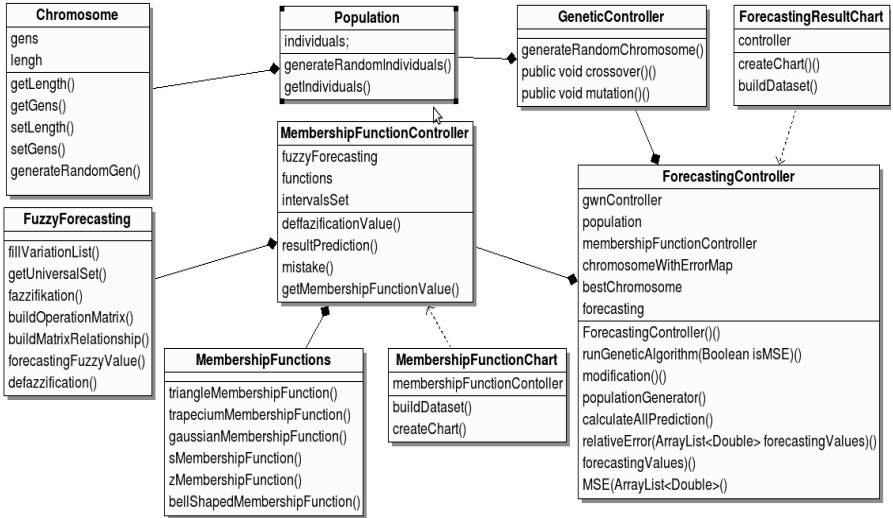


Рис. 1. UML діаграма класів

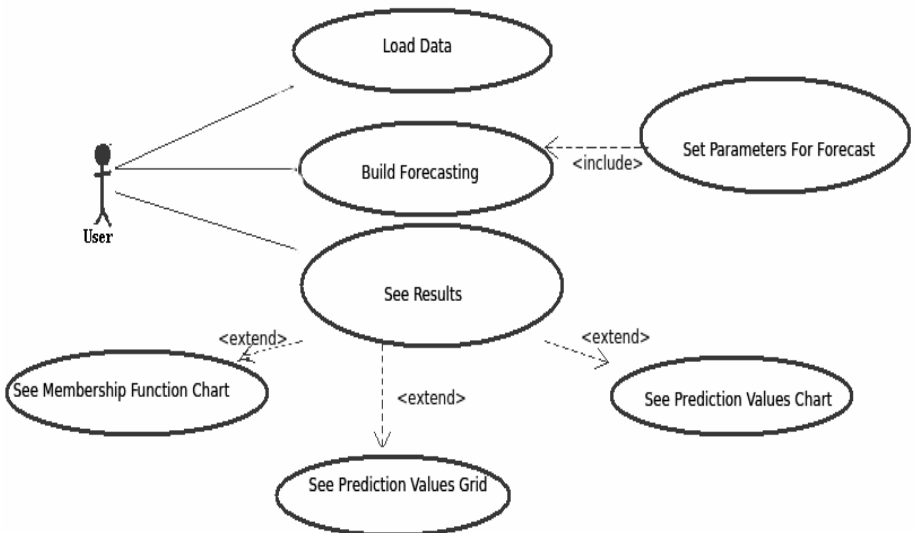


Рис. 2. UseCase діаграма

Розроблене застосування має дружній та зручний інтерфейс для користувачів, дані наочно представлені у вигляді таблиць та графіків.

Апробація розробленого програмного продукту виконана на даних гідрогеохімічного моніторингу території Північного гірничо-збагачувального комбінату (ГЗК) за період з лютого 1986 р. по травень 1992 р. На рис. 3 наведено графік функцій приналежності. На рис. 4 зображено графік прогнозу, побудованого за вище описаним алгоритмом. Історичні дані для прогнозу – це дані концентрації SO_2 у воді. У таблиці 1 наводяться обчислені значення середньоквадратичної та середньої відносної помилки для наборів даних SO_2 , SO_4 , SO , Na , S .

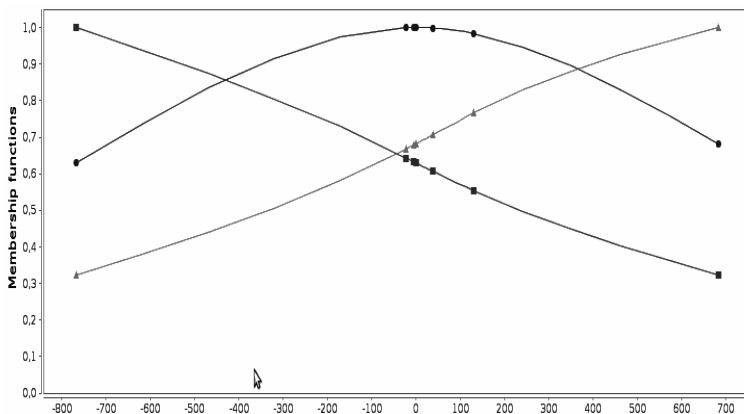


Рис. 3. Графіки функцій приналежності

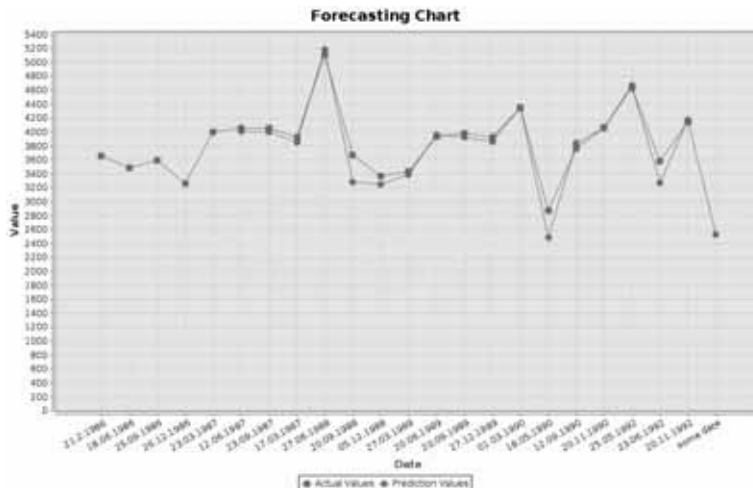


Рис. 4. Графік отриманих результатів

Таблиця 1

Обчислені похибки прогнозу

Дані	Середньоквадратична помилка	Середня відносна помилка (%)
SO ₂	568231	2.77
SO ₄	691091	2.82
SO	1924084	2.40
Na	176672	2.81
S	1682336	2.76

Висновки. У роботі розглянута інформаційна технологія для прогнозування нечітких часових рядів. Розглянута можливість застосування нечіткої логіки для аналізу та прогнозування часових рядів концентрації хімічних компонентів у пробах підземних вод на території гірничо-збагачувального комбінату. Подальші дослідження можуть бути спрямовані на вдосконалення розглянутого підходу для виявлення більш складних шаблонів у нечітких часових рядах.

Бібліографічні посилання

1. **Леоненков А.В.** Нечеткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. – СПб., 2005. – 736 с.
2. **Мамедова М.Г.** Нечеткая логика в прогнозировании демографических аспектов рынка труда / М.Г. Мамедова, З.Г. Джабраилова // Сб. Трудов НИУЦ по труду и социальным проблемам. – 2005.
3. **Хмелевой С.В.** Инструментальные средства для создания базы знаний на основе нечетких продукций, настраиваемых с помощью генетических алгоритмов./ С.В. Хмелевой, Ю.А. Скобцов, А.М. Фонотов //Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: Матер. II Всеукр. науково-практ. конф. 8–10 квітня 2008 р. – Луганськ: Альма-матер. – 2008. – С.103–105.
4. **Shyi-Ming Chen.** Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms / Shyi-Ming Chen, Nien-Yi Chung // Information and Management Sciences. – Vol. 17, № 3. 2006.
5. **Song Q.** Fuzzy time series and its models. / Q. Song, B.S. Chissom B.S. // Fuzzy Sets and Syst. – № 54, – 1993.
6. **Song Q.** Forecasting enrollments with fuzzy time series – part 1./ Q. Song, B.S. Chissom // Fuzzy Sets and Syst. – № 54, – 1993.
7. **Song Q.** Forecasting enrollments with fuzzy time series – part 2./ Q. Song, B.S. Chissom B.S. // Fuzzy Sets and Syst. – № 54, – 1993.
8. **Chen S.-M.** Forecasting Enrollments Based on Fuzzy Time Series // Fuzzy Sets and Syst. – 81.1996.
9. **Şah, M.** Forecasting Enrollment Model Based on First-Order Fuzzy Time Series. / Degtiarev, K.Y. // Proc. Int. Conf. Computational Intelligence. – 2004.

10. **Shyi-Ming** Chen. Forecasting Enrollments of Students by Using Fuzzy Time Series and Genetic Algorithms / Nien-Yi Chung // Information and Management Sciences. – Vol. 17. – Number 3. – 2006. – P. 1–7.

11. **Cheng, Ching-Hsue**. Predicting daily ozone concentration maxima using fuzzy time series based on a two-stage linguistic partition method / Ching-Hsue Cheng, Sue-Fen Huang, Hia-Jong Teoh // Computers & Mathematics with Applications. – Vol. 62, Issue 4. – 2011. – P. 2016–2028.

12. **Zounemat-Kermani**. Using adaptive neuro-fuzzy inference system for hydrological time series prediction / Mohammad Teshnehlab // Applied Soft Computing. – Vol. 8, Issue 2. – 2008. – P. 928–936.

13. **S. Marsili-Libelli**. Fuzzy pattern recognition of circadian cycles in ecosystems / S. Marsili-Libelli // Ecological Modelling. – Vol. 174, Issues 1–2. – 2004. – P. 67–84.

14. **Elisabetta Giusti**. Spatio-temporal dissolved oxygen dynamics in the Orbetello lagoon by fuzzy pattern recognition / Elisabetta Giusti, Stefano Marsili-Libelli // Ecological Modelling. – Vol. 220, Issue 19. – 2009. – P. 2415–2426.

15. **Alexandra P.** Development of rainfall–runoff models using Takagi–Sugeno fuzzy inference systems / Alexandra P. Jacquin, Asaad Y. Shamseldin // Journal of Hydrology. – Vol. 329, Issues 1–2. – 2006. – P. 154–173.

Надійшла до редколегії 09.07.2012