



О. Ю. Мулеца¹, Ф. Е. Гече¹, А. Є. Батюк², О. О. Мельник¹

¹ Ужгородський національний університет, м. Ужгород, Україна

² Національний університет "Львівська політехніка", м. Львів, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ МЕТОДОМ СИНТЕЗУ ПРОГНОЗНОЇ СХЕМИ

Розроблено інформаційну технологію прогнозування на підставі часових рядів. З'ясовано, що актуальною є розробка нових моделей і методів прогнозування для покращення якості прогнозу. В основу інформаційної технології покладено еволюційний метод синтезу прогнозованої схеми на підставі базових прогнозних моделей. Обраний метод дає змогу вносити в розгляд будь-яку кількість прогнозних моделей, які можуть належати до різних класів. Для заданого часового ряду, шляхом знаходження розв'язку оптимізаційної задачі, обчислено вагові коефіцієнти, з якими моделі входять в результуючу прогнозувану схему. Показано спосіб побудови цільової функції для задачі оптимізації у вигляді лінійної комбінації результатів прогнозування базовими прогнозними моделями. Запропоновано розв'язок оптимізаційної задачі знаходити за допомогою генетичного алгоритму. Результатом роботи методу є прогнозна схема, яка є лінійною комбінацією базових прогнозних моделей. Для оцінювання якості прогнозу запропоновано застосовувати похибки прогнозування або волатильність прогнозу, яка обчислено у вигляді середньоквадратичного відхилення. Критерії якості прогнозу обрано залежно від контексту задачі. Використання волатильності прогнозу як критерію якості, після багаторазового використання технології, дасть змогу зменшити відхилення прогнозних значень від реальних даних. Розроблено структурну схему інформаційної технології. Структурно інформаційна технологія складається з двох блоків: оброблення даних, інтерпретації отриманих значень. Результатом застосування розробленої інформаційної технології є продукційні правила для визначення прогнозного значення досліджуваної величини. Виконано експериментальну верифікацію отриманих результатів. Розв'язано задачу прогнозування кількості релігійних організацій в Україні на підставі статистичних даних з 1997 по 2000 роки. Як базові прогнозні моделі було обрано метод авторегресії та лінійну регресійну модель. За результатами використання розробленої інформаційної технології було обчислено вагові коефіцієнти базових моделей. Показано, що отримана прогнозна схема дала змогу покращити середню абсолютну відсоткову похибку та волатильність прогнозу, порівняно з обраними моделями.

Ключові слова: інформаційна технологія; часовий ряд; прогнозування; еволюційні технології; волатильність прогнозу; синтез прогнозованої схеми.

Вступ

Прийняття рішень у різних сферах людської діяльності пов'язані з аналізом кількісних характеристик відповідних процесів та явищ. Визначальним етапом у таких процесах є визначення прогнозних значень ключових показників. Від точності та якості прогнозу залежить ефективність прийнятих управлінських рішень. Серед задач прогнозування вагоме місце займає задача прогнозування на підставі ретроспективних даних, яку зазвичай формулюють у вигляді задачі прогнозування на підставі часових рядів [10]. Відомо багато класів методів прогнозування на підставі часових рядів, які відрізняються між собою за точністю та простотою використання [15], [16]. Різні методи використовують різні математичні прийоми щодо аналізу внутрішніх закономірностей ряду, серед яких статистичні моделі та методи [21], [22], неймережні методи [25], [28], еволюційні [3] тощо. Значного поширення набули комбіновані методи прогнозування [8], [20], які охоплюють декілька базових прогнозних моделей.

Проте, як вже доведено [27], неможливо побудувати одного найкращого методу прогнозування, який би покращував результати застосування будь-якого з відомих методів для кожного часового ряду. Отже, актуальною є розробка нових інструментів для аналізу часових рядів та обчислення прогнозних значень числових характеристик досліджуваних процесів і явищ.

Об'єкт дослідження – інформаційно-аналітичний супровід аналізу даних різної природи.

Предмет дослідження – інформаційні технології прогнозування на підставі часових рядів.

Мета роботи – проектування інформаційної технології прогнозування часових рядів на підставі еволюційного методу синтезу прогнозованої схеми для покращення якості прогнозу порівняно з обраними прогнозними моделями.

Для досягнення зазначеної мети визначено такі *основні завдання дослідження*:

- проаналізувати еволюційний метод синтезу прогнозованої схеми та визначити критерії якості прогнозу;
- побудувати продукційні правила, які стануть основою вибору кращого прогнозу для заданого часового ряду;
- спроектувати структурно-функціональну схему інформаційної технології прогнозування часових рядів;
- виконати експериментальну верифікацію отриманих результатів.

Наукова новизна отриманих результатів дослідження – визначили продукційні правила, які формуються внаслідок використання розробленої інформаційної технології, що дає змогу обчислювати прогнозне значення досліджуваної величини та забезпечувати покращення якості прогнозу порівняно з базовими прогнозними моделями.

Практична значущість результатів дослідження – використання розробленої інформаційної технології дає змогу покращувати якість прогнозу для кожного заданого часового ряду. У випадку, коли критерієм якості прогнозу є волатильність, після багаторазового використання технології, забезпечується зменшення абсолютного значення відхилення прогнозного значення від реальних даних.

Матеріали і методи дослідження. Під час дослідження використовували як загальнонаукові методи, так і системний підхід, методи системного аналізу, методи прогнозування часових рядів тощо [8], [19], [20]. Для верифікації результатів дослідження застосували статистичні дані щодо кількості релігійних організацій на території України з 1997 по 2020 рр.

Аналіз останніх досліджень та публікацій. Інформаційні технології та їхні компоненти, з метою отримання нових даних та знань у різних сферах людської діяльності, розроблено в багатьох сучасних наукових публікаціях [1], [11], [14], [18]. Основою для отримання нових знань є спеціально розроблені моделі та методи оброблення висновків експертів [12], [23], [24]; нейромережні методи аналізу даних [9], [17]; QoS методи [4]; методи факторного аналізу [6], [5] тощо. Методи прогнозування на підставі часових рядів базуються на різних підходах аналізу закономірностей, які є в ряді. У роботі [15] розроблено метод машинного навчання для класифікації часових рядів. У роботі [16] використано фрактальний аналіз часових рядів. У роботі [20] запропоновано метод врахування суджень експертів під час розв'язання задач прогнозування.

Проблемі розроблення гібридних методів прогнозування присвячено ряд наукових публікацій. Наприклад, у роботі [8] розроблено комбіновану схему прогнозування на підставі базових прогнозних моделей, де моделі обрано відповідно до принципу конкуренції. Розробці еволюційного методу побудови прогнозної схеми присвячена [19]. У цих роботах показано, що за допомогою використання комбінованих моделей прогнозування можливим є покращення якості результатів прогнозу, порівняно з базовими методами. Оскільки розв'язання задачі прогнозування зазвичай приводить до розв'язання оптимізаційних задач, то актуальним є розроблення нових інформаційних технологій, в основі яких є сучасні моделі та методи прогнозування на підставі часових рядів.

Результати дослідження та їх обговорення

Еволюційний метод синтезу прогнозної схеми. Розглянемо задачу прогнозування на підставі часових рядів у такій постановці: Нехай $v_1, v_2, \dots, v_i, \dots, v_n$ – деякий дискретний часовий ряд без пропусків довжини n . Значення часового ряду фіксуються у дискретні моменти часу $i = \overline{1, n}$. Задача прогнозування часових рядів полягає у побудові вирішального правила F , яке дає змогу знайти оцінку значення ряду в майбутні періоди часу, що є прогнозом у заданій точці $n+m$ з кроком прогнозу m ($m \geq 1$) [19]. Тоді, формально, залежність прогнозного значення \tilde{v}_{n+m} від ретроспективних даних можна записати:

$$\tilde{v}_{n+m} = F(v_{n-k+1}, v_{n-k+2}, \dots, v_n, m),$$

де: k – глибина передісторії, m – крок прогнозування.

Розглянемо еволюційний метод прогнозування на підставі базових моделей прогнозування [19]. Для цього на першому етапі виберемо систему базових моделей прогнозування M_1, M_2, \dots, M_q . Серед них можуть бути регресійні, нейромережні та інші методи прогнозування.

Зафіксуємо значення кроку прогнозування m ($m \geq 1$) та глибини передісторії k . Обчислимо відповідні прогнозні значення для відомих елементів ряду обраними базовими методами та запишемо їх у таблицю (табл. 1):

Табл. 1. Результати прогнозування базовими прогнозними моделями

Номер елемента	Прогнозні моделі			
	M_1	M_2	...	M_q
$k+m$	$\tilde{v}_{k+m}^{(1)}$	$\tilde{v}_{k+m}^{(2)}$...	$\tilde{v}_{k+m}^{(q)}$
$k+m+1$	$\tilde{v}_{k+m+1}^{(1)}$	$\tilde{v}_{k+m+1}^{(2)}$...	$\tilde{v}_{k+m+1}^{(q)}$
...
n	$\tilde{v}_n^{(1)}$	$\tilde{v}_n^{(2)}$...	$\tilde{v}_n^{(q)}$

Побудуємо прогнозу схему у вигляді функції для обчислення прогнозного значення часового ряду в момент часу p ($k+m < p \leq n+1$) за правилом:

$$G(p) = \alpha_1 \tilde{v}_p^{(1)} + \alpha_2 \tilde{v}_p^{(2)} + \dots + \alpha_q \tilde{v}_p^{(q)}, \quad (1)$$

де α_j ($j = \overline{1, q}$) – ваговий коефіцієнт прогнозної моделі M_j для обчислення прогнозного значення часового ряду в момент часу p ($k+m < p \leq n+1$).

Для обчислення значень параметрів схеми α_j ($j = \overline{1, q}$) при заданому значенні p побудуємо функціонал

$$L(\alpha_1, \alpha_2, \dots, \alpha_q) = \sum_{i=k+m}^{p-1} \left| \frac{G(i) - v_i}{v_i} \right|, \quad (2)$$

та застосуємо генетичний алгоритм [26] для розв'язання задачі мінімізації функції

$$L(\alpha_1, \alpha_2, \dots, \alpha_q) \rightarrow \min. \quad (3)$$

Прогнозне значення \tilde{v}_p можна обчислити за формулою

$$\tilde{v}_p = \alpha_1^* \tilde{v}_p^{(1)} + \alpha_2^* \tilde{v}_p^{(2)} + \dots + \alpha_q^* \tilde{v}_p^{(q)}, \quad (4)$$

де $(\alpha_1^*, \alpha_2^*, \dots, \alpha_q^*)$ – розв'язок задачі (3), тобто $(\alpha_1^*, \alpha_2^*, \dots, \alpha_q^*) \in \text{Arg min } L(\alpha_1, \alpha_2, \dots, \alpha_q)$

Умови якості прогнозної схеми. У разі застосування еволюційного методу синтезу прогнозної схеми важливою є проблема оцінювання якості отриманих результатів прогнозу. Залежно від особливостей процесу чи явища, кількісні показники якого відображені в часовому ряді, мають місце різні підходи до її вирішення. Найпоширенішим способом оцінки якості прогнозу є розрахунок похибок прогнозу, серед яких найбільш уживаними є [2], [10], [13]:

- середнє абсолютне відхилення

$$MAD = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} |v_i - \tilde{v}_i|; \quad (5)$$

- середня абсолютна відсоткова похибка

$$MAPE = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} \frac{|v_i - \tilde{v}_i|}{v_i}; \quad (6)$$

- середня відсоткова похибка

$$MPE = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} \frac{v_i - \tilde{v}_i}{v_i}, \text{ тощо,} \quad (7)$$

де: m – крок прогнозування, $m \geq 1$, k – глибина передісторії, p – номер елемента ряду, для прогнозування якого будується прогнозна схема, \tilde{v}_i , $i = p_0, p-1$, – прогнозні значення, обчислені за (4), p_0 – фіксований параметр, $k + m \leq p_0 < p$.

Вважатимемо, що прогнозна схема (4) покращує результати прогнозу, отримані за допомогою прогнозуючих моделей M_1, M_2, \dots, M_q , якщо для даних, обчислених за допомогою схеми, забезпечується зменшення похибки прогнозу. Для виконання такого порівняння обчислюємо значення похибок прогнозу за вибраним правилом в один з таких способів:

$$MAD^{(j)} = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} |v_i - \tilde{v}_i^{(j)}|, j = \overline{1, q},$$

$$MAPE^{(j)} = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} \frac{|v_i - \tilde{v}_i^{(j)}|}{v_i}, j = \overline{1, q},$$

$$MPE^{(j)} = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} \frac{v_i - \tilde{v}_i^{(j)}}{v_i}, j = \overline{1, q}.$$

Тоді якщо похибка прогнозу синтезованої прогновної схеми менша за похибки прогнозу, отримані внаслідок використання прогнозних моделей, тобто, наприклад, виконуються умови, що

$$\forall j = \overline{1, q} \quad MAPE \leq MAPE^{(j)}$$

та $\exists j_0 \in \{1, 2, \dots, q\} : MAPE < MAPE^{(j_0)}$,

то будемо вважати, що прогнозна схема (4) дає змогу отримати кращі результати прогнозу порівняно з обраними прогнозними моделями M_1, M_2, \dots, M_q .

Використання цих похибок дає змогу обчислити середнє значення відхилення результатів прогнозу від реальних даних. Проте, за мінімізації одного із значень (5)–(7), можливим є випадок, при якому

$$\exists i_0 \in \{p_0, p_0 + 1, \dots, p - 1\} : |v_{i_0} - \tilde{v}_{i_0}| > \varepsilon, \quad (8)$$

де $\varepsilon > 0$ – заданий поріг, який показує на скільки прогнозні значення можуть відхилитися від реальних даних. Виконання умови (8) означає, що для деякого елемента часового ряду прогнозне значення відхиляється від реального більше, ніж це допустимо з огляду на контекст задачі прогнозування.

Позначимо через $\varepsilon_{\max} = \max_{i=p_0, p-1} |v_i - \tilde{v}_i|$. Тоді можливим є випадок, в якому є потреба розв'язання задачі (9)

$$\varepsilon_{\max} \rightarrow \min, \quad (9)$$

тобто знаходження таких параметрів прогновної схеми (4), які б дали змогу зменшити відхилення прогнозних значень від реальних даних у разі багаторазового розв'язання задачі прогнозування. Для розв'язання задачі (9), у процесі прогнозування на підставі часових рядів, пропонуємо розглядати волатильність прогнозу як критерій якості прогновної схеми.

Для оцінювання волатильності отриманого прогнозу, пропонуємо використовувати середньоквадратичне відхилення [19]. Обчислимо

$$\sigma_j = \sqrt{\frac{1}{p - p_0} \sum_{i=p_0}^{p-1} (\tilde{v}_i^{(j)} - v_i - \delta_j)^2}, j = \overline{1, q}, \quad (10)$$

$$\sigma_0 = \sqrt{\frac{1}{p - p_0} \sum_{i=p_0}^{p-1} (G(i) - v_i - \delta_0)^2}, \quad (11)$$

$$\text{де } \delta_j = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} (\tilde{v}_i^{(j)} - v_i), j = \overline{1, q}, \delta_0 = \frac{1}{p - p_0} \sum_{i=p_0}^{p-1} (G(i) - v_i).$$

Тоді якщо виконуються умови $\forall j = \overline{1, q} \quad \sigma_0 \leq \sigma_j$ та $\exists j_0 \in \{1, 2, \dots, q\} : \sigma_0 < \sigma_{j_0}$, то вважатимемо, що прогнозна схема (4) є якісною порівняно з базовими прогнозуючими моделями.

Використання волатильності як критерію якості прогнозу є актуальним під час прогнозування в медицині, охороні здоров'я, економіці тощо. У таких задачах, у разі багаторазового прогнозування, важливішим є забезпечення мінімізації дисперсії, ніж точності прогнозу на сьогодні.

Проектування інформаційної технології прогнозування часових рядів. Інформаційна технологія прогнозування часових рядів на підставі еволюційного методу синтезу прогновної схеми є сукупністю методів та засобів нагромадження, оброблення і зберігання даних з метою отримання нових знань щодо значень прогнозованої величини. Вхідними даними є часові ряди, основу технології становлять моделі та методи прогнозування. Результатом застосування інформаційної технології є правило щодо обчислення прогнозних значення досліджуваної величини. Структурну схему розробленої інформаційної технології навели на рисунку.



Рисунок. Структурна схема інформаційної технології

Відповідно до рисунку, часовий ряд подається на етап оброблення даних, де виконується прогнозування на підставі обраних прогнозних моделей та синтезується прогнозна схема (4). Етап інтерпретації даних полягає в обчисленні похибок прогнозу (5)–(7) та волатильності (10)–(11) отриманих результатів.

Продукційні правила, які формуються внаслідок використання інформаційної технології, у випадку, коли критерієм прогнозу є волатильність, є такими:

- якщо $\forall j = \overline{1, q} \quad \sigma_0 \leq \sigma_j$ і $\exists j_0 \in \{1, 2, \dots, q\} : \sigma_0 < \sigma_{j_0}$, то $\tilde{v}_p = \alpha_1^* \tilde{v}_p^{(1)} + \alpha_2^* \tilde{v}_p^{(2)} + \dots + \alpha_q^* \tilde{v}_p^{(q)}$;
- якщо $\exists j_0 \in \{1, 2, \dots, q\} : \sigma_{j_0} < \sigma_0$, то $\tilde{v}_p = \tilde{v}_p^{(j_0)}$;
- якщо $\min_{j=\overline{1, q}} \sigma_j = \sigma_0$, то $\tilde{v}_p = \tilde{v}_p^{(j)}$,

де: $j' \in \text{Arg} \min_{j=\overline{1, q}} \sigma_j, j' \in M$; $M = \text{Arg} \min_{j=\overline{1, q}} \sigma_j \cup \{0\}$;

$$MAPE^{(j)} = \min_{j \in M} MAPE^{(j)}.$$

Залежно від контексту задачі, продукційні правила можуть бути відкориговані.

Експериментальна верифікація отриманих результатів. Для експериментальної верифікації результатів дослідження розв'язали задачу прогнозування кількості релігійних організацій в Україні за даними Релігійно-інформаційної служби України [7]. Як базові прогнозні моделі обрали метод авторегресії (AR (2)) та лінійну регресійну модель з кроком передісторії $k=5$ та кроком прогнозу $m=1$.

Для покращення якості прогнозу на наступному етапі застосували еволюційний метод синтезу прогнозної схеми з цільовою функцією такого виду:

$$L_p(\alpha_1, \alpha_2) = \sum_{i=6}^{p-1} \left| \frac{\alpha_1 \tilde{v}_i^{(1)} + \alpha_2 \tilde{v}_i^{(2)} - v_i}{v_i} \right|,$$

де: $\tilde{v}_i^{(1)}$ – результат прогнозування i -го значення ряду методом авторегресії, $\tilde{v}_i^{(2)}$ – результат прогнозування i -го значення ряду за допомогою лінійної регресійної моделі, $p = 10, 21$. Прогнозне значення \tilde{v}_p обчислювали за такою формулою

$$\tilde{v}_p = \alpha_1^{*(p)} \tilde{v}_p^{(1)} + \alpha_2^{*(p)} \tilde{v}_p^{(2)},$$

де $(\alpha_1^{*(p)}, \alpha_2^{*(p)}) = \arg \min L_p(\alpha_1, \alpha_2)$, $p = 9, 21$.

Для оцінки якості отриманого прогнозу різними методами обчислювали похибки $MAPE$ за правилом (6) та волатильність (10)-(11). Отримані результати наведено в табл. 2.

Табл. 2. Порівняння результатів прогнозування значень часового ряду "Релігійні організації", прогнозною схемою та базовими прогнозними моделями

Рік	v_i	$\tilde{v}_i^{(1)}$	$\tilde{v}_i^{(2)}$	$G(i)$		
				α_1^*	α_2^*	\tilde{v}_i
1997	17610	-	-	-	-	-
1998	19005	-	-	-	-	-
1999	20309	-	-	-	-	-
2000	21693	-	-	-	-	-
2001	23400	-	-	-	-	-
2002	24919	25225	24683,8	-	-	-
2003	26271	26497	26341,4	-	-	-
2004	27579	27582	27863,0	-	-	-
2005	28481	28741	29165,8	-	-	-
2006	29262	29151	29976,2	0,740179	0,254960450	29219,73
2007	29913	29832	30571,6	0,302420	0,687121168	30028,22
2008	30670	30331	30991,2	0,318367	0,672312752	30492,19
2009	31257	31370	31465,4	0,429340	0,564741974	31238,23
2010	31257	31768	32005,0	0,800513	0,193951105	31638,12
2011	32521	31043	32072,4	0,622787	0,371884149	31260,41
2012	32521	33059	32864,8	0,529296	0,467895772	32875,31
2013	33581	32411	33134,6	0,048660	0,941323865	32767,52
2014	35646	33812	34001,2	0,323340	0,674391419	33862,90
2015	33781	37039	36056,8	0,151498	0,851816105	36325,10
2016	34183	33853	35304,0	0,525619	0,474036675	34529,19
2017	34385	33966	34999,4	0,572245	0,421278370	34181,40
2018	34637	34151	34359,0	0,492930	0,501527284	34066,04
2019	35162	34896	34102,6	0,586335	0,406030048	34307,34
2020	36796	35645	35394,6	0,898710	0,097159491	35473,57
MAPE		0,023797405	0,024371272	0,021058799		
σ		1138,503501	988,33730	981,7719415		

Як видно з табл. 2, хоча точково, прогнозна схема не завжди покращує результати прогнозування базовими прогнозними моделями, проте за її використання можливим є покращення як похибок прогнозування, так і волатильності прогнозу. Це дає змогу під час багаторазового використання методу, в кожному конкретному випадку, забезпечити зменшення відхилення прогнозних значень від реальних даних порівняно з базовими прогнозними моделями.

Результати прогнозу на 2021 р. наведено в табл. 3.

Табл. 3. Прогнозні значення часового ряду "Релігійні організації"

Рік	$\tilde{v}_i^{(1)}$	$\tilde{v}_i^{(2)}$	$G(i)$		
			α_1^*	α_2^*	\tilde{v}_i
2021	38560	36833,8	0,662930289	0,331388994	37768,91

Отже, прогнозне значення на 2021 р. – близько 37769 організацій.

Обговорення результатів дослідження. Розробили інформаційну технологію прогнозування, в основу якої поставили еволюційний метод синтезу прогнозної схеми. Особливістю розробленої інформаційної технології є те, що її можна наповнювати різними прогнозними моделями, які стануть базовими для запропонованого еволюційного методу прогнозування на підставі часових рядів. Покращення якості прогнозу забезпечується шляхом мінімізації побудованої оптимізаційної функції, в яку результати прогнозування базовими моделями входять з відповідними ваговими коефіцієнтами. Вагові коефіцієнти є аргументами цільової функції.

У результаті експериментальної верифікації досягнули покращення якості прогнозу шляхом використання розробленої інформаційної технології. Порівняння здійснили методом авторегресії та лінійною регресійною моделлю. Значення середньої абсолютної від-

соткової похибки зменшили від 0,023797405 та 0,024371272 до 0,021058799; середньоквадратичне відхилення – від 1138,503501 і 988,3373 до 981,7719415.

Перспективними є дослідження, які пов'язані із залученням у розроблену інформаційну технологію експертів для можливості уточнення прогнозних значень на підставі їхніх висновків.

Висновок

1. Проаналізовано еволюційний метод синтезу прогнозної схеми на підставі базових прогнозних моделей. Визначено, що критеріями якості прогнозу, залежно від мети та контексту задачі прогнозування, можуть бути похибки прогнозування або волатильність прогнозу. Використання різних підходів до визначення якості прогнозу дає змогу як зменшити середню похибку прогнозування, так і забезпечити мінімізацію абсолютного значення відхилення прогнозних значень від реальних даних.

2. Побудовано продукційні правила вибору кращого за обраним критерієм якості прогнозного значення для заданого часового ряду. Параметри продукційних правил формуються внаслідок використання розробленої інформаційної технології. Результатом використання правил є числове значення прогнозованої величини.

3. Спроектовано структурну схему інформаційної технології для прогнозування на підставі часових рядів. Основою технології є еволюційний метод синтезу прогнозної схеми. Передбачено можливість наповнення схеми прогнозними моделями різних видів. Результатом використання технології є нові знання у формі продукційних правил, які використовуються для автоматизованого визначення шуканого прогнозного значення.

4. Виконано експериментальну верифікацію результатів дослідження для прогнозування значень часового ряду "Релігійні організації". Для аналізу було відібрано метод авторегресії та лінійну регресійну модель. Показано, що за допомогою технології було покращено як похибки прогнозу, так і його волатильності. Це дасть змогу під час багаторазового застосування інформаційної технології зменшити відхилення прогнозних значень від реальних даних.

References

- [1] Biloshchytskyi, A., Biloshchytska, S., Kuchansky, A., Bielova, O., & Andrashko, Y. (2018, February). Infocommunication system of scientific activity management on the basis of project-vector methodology. In *2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, 200–203. IEEE. <https://doi.org/10.1109/TCSET.2018.8336186>
- [2] Brockwell, P. J., Brockwell, P. J., Davis, R. A., & Davis, R. A. (2016). *Introduction to time series and forecasting*, Springer, 434. <https://doi.org/10.1007/978-3-319-29854-2>
- [3] Cai, Q., Zhang, D., Zheng, W., & Leung, S. C. (2015). A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, 74, 61–68. <https://doi.org/10.1016/j.knosys.2014.11.003>
- [4] Daradkeh, Y. I., Kirichenko, L., & Radivilova, T. (2018). Development of QoS methods in the information networks with fractal traffic. *International Journal of Electronics and Telecommunications*, 64, 27–32.
- [5] Dolgikh, S. (2019). Categorized representations and general learning. In *International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions*, August 2019, Springer, Cham, 93–100. https://doi.org/10.1007/978-3-030-35249-3_11
- [6] Dolgikh, S., & Mulesa, O. (2021) Covid-19 epidemiological factor analysis: Identifying principal factors with machine learning. CEUR Workshop Proceedings, 2833, 114–123. <https://doi.org/10.1101/2020.06.01.20119560>
- [7] Ethnopolitics. (2020). State Service of Ukraine for Ethnopolitics and Freedom of Conscience. Retrieved from: <https://dcss.gov.ua/statistics-2020/>
- [8] Geche, F., Batyuk, A., Mulesa, O., & Voloshchuk, V. (2020, August). The Combined Time Series Forecasting Model. In *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, 272–275, IEEE. <https://doi.org/10.1109/DSMP47368.2020.9204311>
- [9] Geche, F., Mulesa, O., & Buchok, V. (2017). Synthesis of generalized neural elements by means of the tolerance matrices. *Eastern-European Journal of Enterprise Technologies*, 4(4), 50–62. <https://doi.org/10.15587/1729-4061.2017.108404>
- [10] Gulyanitsky, L. F., & Bondar, T. G. (2018). Research of efficiency of adaptive forecasting methods. *Computer Mathematics*, (1), 53–60.
- [11] Hnatiienko, H., Kudin, V., Onyshchenko, A., Snytyuk, V., & Kruhlov, A. (2020, October). Greenhouse Gas Emission Determination Based on the Pseudo-Base Matrix Method for Environmental Pollution Quotas Between Countries Allocation Problem. In *2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC)*, 1–8. IEEE. <https://doi.org/10.1109/SAIC51296.2020.9239125>
- [12] Hnatiienko, H., Tmienova, N., & Kruglov, A. (2021) Methods for Determining the Group Ranking of Alternatives for Incomplete Expert Rankings. In: Shkarlet S., Morozov A., Palagin A. (eds) *Mathematical Modeling and Simulation of Systems (MODS2020)*. MODS 2020. *Advances in Intelligent Systems and Computing*, 1265, 217–226, Springer, Cham. https://doi.org/10.1007/978-3-030-58124-4_21
- [13] Hunk, D., Rights, A. D., & Dean, W. (2003). *Business forecasting*, Williams.
- [14] Khomytska, I., Teslyuk, V., Kryvinska, N., & Bazylevych, I. (2020). Software-based approach towards automated authorship acknowledgement – Chi-square test on one consonant group. *Electronics*, 9(7). <https://doi.org/10.3390/electronics9071138>
- [15] Kirichenko, L., Radivilova, T., Bulakh, V., Zinchenko, P., & Alghawli, A. S. (2020, August). Two Approaches to Machine Learning Classification of Time Series Based on Recurrence Plots. In *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, August 2020, IEEE, 84–89. <https://doi.org/10.1109/DSMP47368.2020.9204021>
- [16] Kuchansky, A., Biloshchytskyi, A., Bronin, S., Biloshchytska, S., & Andrashko, Y. (2019, October). Use of the Fractal Analysis of Non-stationary Time Series in Mobile Foreign Exchange Trading for M-Learning. In *Interactive Mobile Communication, Technologies and Learning* (pp. 950–961). Springer, Cham. https://doi.org/10.1007/978-3-030-49932-7_88
- [17] Lupei, M., Mitsa, A., Povkhan, I., & Sharkan, V. (2020). Determining the Eligibility of Candidates for a Vacancy Using Artificial Neural Networks. In *2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, August 2020, IEEE, 18–22. <https://doi.org/10.1109/DSMP47368.2020.9204020>
- [18] Mulesa, O. Yu. (2015). Design features of the information technology for the labor migrants group structure determination. *Eastern-European Journal of Enterprise Technologies*, 4(2), 4–8. <https://doi.org/10.15587/1729-4061.2015.47204>
- [19] Mulesa, O. Yu., & Snytyuk, V. Ye. (2020). Development of the evolutive method for forecasting hourly rows. *Automation of technological and business processes*, 12(3), 4–9. <https://doi.org/10.15673/atbp.v12i3.1854>
- [20] Mulesa, O., Geche, F., Voloshchuk, V., Buchok, V., & Batyuk, A. (2017). Information technology for time series fore-

- casting with considering fuzzy expert evaluations. In *2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, 1, 105–108, IEEE, September 2017. <https://doi.org/10.1109/STC-CSIT.2017.8098747>
- [21] Pole, A., West, M., & Harrison, J. (2018). *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315274775>
- [22] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- [23] Tsyganok, V. V., Kadenko, S. V., & Andriichuk, O. V. (2011). Simulation of expert judgements for testing the methods of information processing in decision-making support systems. *Journal of Automation and Information Sciences*, 43(12). <https://doi.org/10.1615/JAutomatInfScien.v43.i12.30>
- [24] Tsyganok, V., Kadenko, S., Andriichuk, O., & Roik, P. (2018, October). Combinatorial method for aggregation of incomplete group judgments. In *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SA-IC)*, 1–6. IEEE. <https://doi.org/10.1109/SAIC.2018.8516768>
- [25] Wan, R., Mei, S., Wang, J., Liu, M., & Yang, F. (2019). Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics*, 8(8), 876. <https://doi.org/10.3390/electronics8080876>
- [26] Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65–85. <https://doi.org/10.1007/BF00175354>
- [27] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>
- [28] Yan, W. (2012). Toward automatic time-series forecasting using neural networks. *IEEE transactions on neural networks and learning systems*, 23(7), 1028–1039. <https://doi.org/10.1109/TNNLS.2012.2198074>

O. Yu. Mulesa¹, F. E. Geche¹, A. Ye. Batyuk², O. O. Melnyk¹

¹Uzhhorod National University, Uzhhorod, Ukraine

²Lviv Polytechnic National University, Lviv, Ukraine

INFORMATION TECHNOLOGY FOR TIME SERIES FORECASTING BY THE METHOD OF THE FORECAST SCHEME SYNTHESIS

The study is devoted to the development of information technology for forecasting based on time series. It has been found that it is important to develop new models and forecasting methods to improve the quality of the forecast. Information technology is based on the evolutionary method of synthesis of the forecast scheme grounded on basic forecast models. The selected method allows you to consider any number of predictive models that may belong to different classes. For a given time series, the weight coefficients with which the models are included in the resulting forecast scheme are calculated by finding the solution to the optimization problem. The method of constructing the objective function for the optimization problem in the form of a linear combination of forecasting results by basic forecasting models is shown. It is proposed to find the solution to the optimization problem using a genetic algorithm. The result of the method is the forecast scheme, which is a linear combination of basic forecast models. To assess the quality of the forecast, it is suggested to use forecasting errors or forecast volatility calculated as the standard deviation. Forecast quality criteria are selected depending on the context of the task. The use of forecast volatility as a quality criterion, with repeated use of technology, will reduce the deviation of forecast values from real data. The structural scheme of information technology is developed. Structurally, information technology consists of two blocks: data processing and interpretation of the obtained values. The result of the application of the developed information technology is the production rules for determining the predicted value of the studied quantity. Experimental verification of the obtained results was performed. The problem of forecasting the number of religious organizations in Ukraine based on statistical data from 1997 to 2000 has been solved. The autoregression method and the linear regression model were chosen as the basic forecast models. Based on the results of using the developed information technology, the weights of the basic models were calculated. It is demonstrated that the obtained forecast scheme allowed to improve the average absolute percentage error and forecast volatility in comparison with the selected models.

Keywords: information technology; time series; forecasting; evolutionary technologies; forecast volatility; synthesis of the forecast scheme.

Інформація про авторів:

Мулеса Оксана Юрїївна, д-р тех. наук, професор, кафедра програмного забезпечення систем.

Email: oksana.mulesa@uzhnu.edu.ua; <https://orcid.org/0000-0002-6117-5846>

Гече Федір Елемирович, д-р тех. наук, професор, кафедра фізико-математичних дисциплін. Email: fedir.geche@uzhnu.edu.ua;

<https://orcid.org/0000-0002-4757-9828>

Батюк Анатолій Євгенович, канд. техн. наук, доцент, кафедра автоматизованих систем управління. Email: abatyuk@gmail.com;

<https://orcid.org/0000-0001-7650-7383>

Мельник Олена Олексіївна, канд. техн. наук, доцент, кафедра програмного забезпечення систем.

Email: olena.melnyk@uzhnu.edu.ua; <https://orcid.org/0000-0001-7340-8451>

Цитування за ДСТУ: Мулеса О. Ю., Гече Ф. Е., Батюк А. Є., Мельник О. О. Інформаційна технологія прогнозування часових рядів методом синтезу прогнозуної схеми. *Український журнал інформаційних технологій*. 2021, т. 3, № 2. С. 81–86.

Citation APA: Mulesa, O. Yu., Geche, F. E., Batyuk, A. Ye., & Melnyk, O. O. (2021). Information technology for time series forecasting by the method of the forecast scheme synthesis. *Ukrainian Journal of Information Technology*, 3(2), 81–86.

<https://doi.org/10.23939/ujit2021.02.081>