ISSN 2707-1898 (print)

Український журнал інформаційних технологій
Ukrainian Journal of Information Technology
https://science.lpnu.ua/uk/ujit
https://doi.org/10.23939/ujit2021.03.044
Article received 17.04.2021 p.
Article accepted 01.06.2021 p.
UDK 004.94

Correspondence author
A. Ye. Batyuk
abatyuk@gmail.com

Ф. Е. Гече[1], О. Ю. Мулеса[1], А. Є. Батюк[2], В. Ю. Смоланка[1]

[1] *Ужгородський національний університет, м. Ужгород, Україна*
[2] *Національний університет "Львівська політехніка", м. Львів, Україна*

НАВЧАННЯ КОМБІНОВАНОЇ МОДЕЛІ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Розроблено метод побудови комбінованої моделі прогнозування часових рядів на підставі базових моделей прогнозування. Множина базових моделей є динамічною, тобто у цю множину можуть вноситися нові моделі прогнозування, можуть видалятися моделі залежно від властивостей часових рядів. Для синтезу комбінованої моделі прогнозування часових рядів з заданим кроком прогнозу на початку визначається оптимальний крок передісторії. Будується функціонал і для фіксованого кроку прогнозу методом авторегресії визначається оптимальний крок передісторії, що визначає проміжок часу на якому проводиться аналіз точності моделей з базової множини. У процесі побудови комбінованої моделі для кожної базової моделі визначається ваговий коефіцієнт з яким вона входить у комбіновану модель. Вагові коефіцієнти базових моделей визначаються на підставі їх точності прогнозування на часовому періоді, визначеного кроком передісторії. Вагові коефіцієнти відображають міру впливу базових моделей на точність прогнозування комбінованої моделі. Після побудови комбінованої моделі проводиться її навчання та визначаються ті базові моделі, які будуть внесені в остаточну комбіновану модель прогнозування. Встановлено правило внесення базових моделей у комбіновану модель. При внесенні базових моделей у комбіновану модель прогнозування враховуються їх вагові коефіцієнти, які залежать від одного й того ж параметру. Визначається оптимальне значення параметру через мінімізацію заданого функціоналу, що задає середнє квадратичне відхилення між фактичними і прогнозними значеннями часового ряду. Вагові коефіцієнти з оптимальними параметрами ранжуються у порядку не зростання та використовуються на етапі внесення базових моделей у комбіновану модель. Внаслідок такого підходу, як показують конкретні приклади, у багатьох випадках вдалося істотно покращити точність прогнозування комбінованої моделі.

Ключові слова: часовий ряд; функціонал; модель прогнозування; навчання моделі; крок прогнозу; ваговий коефіцієнт.

Вступ

У ринкових умовах ефективне прогнозування і планування виробничих потужностей підприємств неможливе без створення відповідних економіко-математичних моделей, які обумовлюють використання наявних ресурсів у процесі виробництва і надання послуг.

У визначенні стратегій розвитку підприємств важливу роль відіграє обчислення прогнозів економічних показників і чинників. Якщо є достовірна інформація про діяльність підприємства в минулому, то для отримання потрібних прогнозів, можна застосовувати математичні методи. Ці методи залежать від мети й деталізації прогнозних чинників і середовища. Вибір методу і моделі, що прогнозується залежить від властивості (від внутрішньої закономірності та зовнішніх впливових факторів) часового ряду. Наведений в роботі метод навчання комбінованої моделі сприяє у вирішенні проблеми вибору "найкращої" моделі для заданого часового ряду шляхом синтезу різних прогнозуючих моделей в одну комбіновану модуль, що будується на підставі "конкуренції" базових моделей.

Об'єкт дослідження – синтез структур комбінованих моделей прогнозування часових рядів.

Предмет дослідження – методи, моделі та алгоритми побудови комбінованих моделей прогнозування, що забезпечує їх гнучкість застосування до прогнозування широкого класу часових рядів.

Мета роботи – розроблення нового методу навчання комбінованих моделей прогнозування часових рядів, що дасть змогу істотно покращити точність прогнозування комбінованої моделі.

Для досягнення зазначеної мети визначено такі основні *завдання дослідження*: на етапі оброблення часових рядів визначити вагові коефіцієнти всіх моделей із базової множини; враховуючи вагові коефіцієнти і критерії внесення моделей у комбіновану модель, синтезувати остаточну модель прогнозування.

Наукова новизна отриманих результатів дослідження – вперше розроблено метод вибору базових моделей і визначення вагових коефіцієнтів, з якими їх вносять у комбіновану модель для підвищення точності прогнозування.

Практична значущість результатів дослідження – залежно від властивостей часового ряду, комбінована модель, внаслідок її навчання, містить найефективніші моделі прогнозування із множини базових моделей для заданого часового ряду з ваговими коефіцієнтами, які визначають в процесі навчання. Такий підхід забезпечує гнучкість комбінованої моделі і їх застосування до прогнозування широкого класу часових рядів.

Матеріали та методи дослідження. В процесі проведення даного дослідження, поряд з загальнонауковими методами були використані відомі методи прогнозування часових рядів, серед яких регресійні методи [10], авторегресійна модель [2], метод найменших квадратів з параметром і без параметра [7], лінійна модель Брауна, квадратична модель Брауна [9]. Аналіз ефективності розробленої комбінованої моделі прогнозування виконано для задачі прогнозування обсягів пасажирських перевезень залізничним, річковим і автомобільним транспортами в Україні за період 1980-2013.

Аналіз останніх досліджень та публікацій. Проблемам прогнозування кількісних показників різних процесів та явищ на підставі ретроспективних даних, які знайшли своє відображення у часових рядах, присвячено ряд сучасних наукових досліджень. В основі таких досліджень лежать різні концепції. В роботах [1], [4] виконано порівняльний аналіз основних моделей машинного навчання для прогнозування часових рядів серед яких нейромережні та регресійні моделі. Було наведено відмінності між розглянутими методами прогнозування. Дослідженню відомих методів прогнозування на підставі часових рядів присвячено [12]. В роботі наведено основні теоретичні результати та обґрунтування умов застосування таких відомих методів прогнозування як лінійна регресійна модель, модель авторегресії, логістична регресійна модель тощо.

Предметом великої кількості сучасних досліджень є гібридні прогнозуючі моделі, в яких у різний спосіб поєднано декілька методів прогнозування. Застосування таких моделей часто дає змогу покращувати точність прогнозу шляхом врахування особливостей часових рядів. В роботі [3], з метою підвищення точності прогнозування цін на фондовому ринку, запропоновано дві гібридні моделі прогнозування, які поєднують два види емпіричних режимів декомпозиції EMD та LSTM. Експериментально показано, що запропоновані моделі демонструють кращі показники для прогнозованого ряду, в порівнянні з SVM, MLP тощо. У роботі [8] запропоновано гібридну техніку прогнозування, відповідно до якої елементи часового ряду розділяються на лінійні та нелінійні компоненти з використанням дискретної вейвлет-трансформації (DWT).

Далі, прогнозування здійснюється за допомогою моделей ARIMA та ANN. В роботі [5], [6] містять основні положення методу синтезу прогнозної схеми на підставі базових прогнозних моделей, яка дає змогу для кожного конкретного часового ряду в автоматизованому режимі обирати результати прогнозування тими методами, які забезпечують мінімізацію похибки прогнозування. Робота [11] присвячена розробці еволюційного методу для прогнозування часових рядів, застосування якого дає змогу покращити волатильність прогнозу, в порівнянні з моделями прогнозування, які внесені в базову прогнозуючу модель. В роботі [13] для прогнозування пропонується застосовувати динамічну обчислювальну систему нейронних мереж. В роботі [15] гібридна модель поєднує модель лінійної регресії та нелінійну модель DBN.

Проведений аналіз сучасних наукових джерел дає підстави стверджувати, що актуальною є розробка нових моделей прогнозування, які б давали змогу на підставі реалізації ідеї "конкуренції" базових прогнозних моделей отримувати більш точні результати прогнозування конкретних часових рядів.

Результати дослідження та їх обговорення

1. *Побудова комбінованої моделі прогнозування часових рядів.* Нехай $v_1, v_2, \ldots, v_t, \ldots, v_n$ часовий ряд, що відповідає послідовністю значень певного показника, через який, наприклад, визначається ефективність використання основних засобів підприємства, v_t – значення досліджуваного показника у момент часу *t*, а *n* – кількість реалізацій даного показника.

Прогнозне значення $\tilde{v}_t$ показника v_t у момент часу *v* можна записати в такому вигляді:

$$\tilde{v}_t = f\left(a_1, \ldots, a_r, v_{t-1}, \ldots, v_{t-k}, t\right), \tag{1}$$

де $a_1, \ldots, a_r$ – параметри моделі, k – глибина передісторії. Для знаходження параметрів $a_1, \ldots, a_r$, як правило будуємо функціонал

$$L\left(a_1, \ldots, a_r\right) = \sum_{t=1}^{n}\left(v_t - \tilde{v}_t\right)^2 \tag{2}$$

і мінімізація цього функціоналу проводиться методом найменших квадратів. Після знаходження $a_1^*, \ldots, a_r^*$ параметрів $a_1, \ldots, a_r$, які мінімізують функціонал L, отримані оптимальні параметри прогнозування часового ряду v_t для моделі *f*. Залежно від типу функцій *f* маємо різні оптимальні моделі прогнозування часового ряду.

Для побудови прогнозуючої схеми на початку розглянемо метод авторегресії за допомогою якого визначимо оптимальний крок передісторії k_τ^* для даного часового ряду при фіксованому кроці прогнозу τ. В авторегресійній моделі, як вище було підкреслено, значення показнику v_t у момент часу t залежить від $v_{t-\tau}, v_{t-\tau-1}, \ldots, v_{t-\tau-k_\tau+1}$, де τ – крок прогнозу, k_τ – параметр передісторії при фіксованому τ. Прогнозне значення $\tilde{v}_{n+\tau}$ за методом авторегресії знаходиться за наступною моделлю:

$$\tilde{v}_{n+\tau} = a_1^{(\tau)} v_n + a_2^{(\tau)} v_{n-1} + \ldots + a_{k_\tau}^{(\tau)} v_{n-k_\tau+1}. \tag{3}$$

Для визначення оптимальних значень параметрів $a_t^{*(\tau)}$ при фіксованому $\tau = \tau_0$ мінімізуємо функціонал

$$L\left(a_1^{(\tau)}, \ldots, a_{k_\tau}^{(\tau)}\right) = \sum_{t=k_\tau+\tau}^{n}\left(v_t - a_1^{(\tau)} v_{t-\tau} - \ldots - a_{k_\tau}^{(\tau)} v_{t-\tau-k_\tau+1}\right)^2, \tag{4}$$

тобто розв'язуємо систему рівнянь

$$\frac{\partial L}{\partial a_t^{(\tau)}} = 0, \quad t = 1, 2, \ldots, k_\tau. \tag{5}$$

Нехай $a_1^{*(\tau)}, \ldots, a_{k_\tau}^{*(\tau)}$ є розв'язком системи (5). Тоді, згідно з виразом (3), маємо

$$\tilde{v}_t = a_1^{*(\tau)} v_{t-\tau} + a_2^{*(\tau)} v_{t-\tau-1} + \ldots + a_{k_\tau}^{*(\tau)} v_{t-\tau-k_\tau+1}, \tag{6}$$

де $t \ge k_\tau + \tau$. Очевидно, що значення v_t при фіксованому $\tau = \tau_0$ залежить від параметра $k_\tau \left(1 \le k_\tau \le n - \tau\right)$. Для визначення оптимального значення параметра передісторії k_τ при $\tau = \tau_0$ для заданого часового ряду v_t розглянемо величини

$$\delta_1 = \frac{1}{n-\tau}\sum_{t=\tau+1}^{n}\left(v_t - a_1^{*(\tau)} v_{t-\tau}\right)^2,$$

$$\delta_2 = \frac{1}{n-\tau-1}\sum_{t=\tau+2}^{n}\left(v_t - a_1^{*(\tau)} v_{t-\tau} - a_2^{*(\tau)} v_{t-\tau-1}\right)^2,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\delta_{n-\tau} = \left(v_n - a_1^{*(\tau)} v_{n-\tau} - \ldots - a_{n-\tau}^{*(\tau)} v_1\right)^2$$

і знаходимо $\min\left\{\delta_1, \delta_2, \ldots, \delta_{n-\tau}\right\} = \delta_{k_\tau^*}$. Величина k_τ^* визначає оптимальне значення параметра передісторії моделі авторегресії при фіксованому $\tau = \tau_0$.

Після визначення k_τ^* при фіксованому $\tau = \tau_0$ розглянемо різні моделі прогнозування $M_1, M_2, \ldots, M_q$ часового ряду з кроком прогнозу τ на часовому періоді $n - k_\tau^* + 1,\ n - k_\tau^* + 2, \ldots, n$. На підставі результатів прогнозування за наведеними вище методами побудуємо таку таблицю.

Табл. 1. Прогнозні значення показників відносно різних моделей

Моделі прогнозування	Значення досліджуваного економічного показника за період $n - k_\tau^* + 1,\ n - k_\tau^* + 2, \ldots, n$			
	$v_{n-k_\tau^*+1}$	$v_{n-k_\tau^*+2}$	$\ldots$	v_n
M_1	$\tilde{v}^{(1)}_{n-k_\tau^*+1}$	$\tilde{v}^{(1)}_{n-k_\tau^*+2}$	$\ldots$	$\tilde{v}^{(1)}_n$
M_2	$\tilde{v}^{(2)}_{n-k_\tau^*+1}$	$\tilde{v}^{(2)}_{n-k_\tau^*+2}$	$\ldots$	$\tilde{v}^{(2)}_n$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
M_q	$\tilde{v}^{(q)}_{n-k_\tau^*+1}$	$\tilde{v}^{(q)}_{n-k_\tau^*+2}$	$\ldots$	$\tilde{v}^{(q)}_n$

У кожному стовпці $v_{n-k_\tau^*+1}, v_{n-k_\tau^*+2}, \ldots, v_n$ табл. 1 знаходимо найменше квадратичне відхилення між прогнозними і фактичними значеннями відповідних членів часового ряду. Математично це можна записати так. Нехай

$$j_1 = n - k_\tau^* + 1 \text{ і}$$

$$\varepsilon_1 = \min\left\{\left(v_{j_1} - \tilde{v}^{(1)}_{j_1}\right)^2, \left(v_{j_1} - \tilde{v}^{(2)}_{j_1}\right)^2, \ldots, \left(v_{j_1} - \tilde{v}^{(q)}_{j_1}\right)^2\right\},$$

$$j_2 = n - k_\tau^* + 2 \text{ і}$$

$$\varepsilon_2 = \min\left\{\left(v_{j_2} - \tilde{v}^{(1)}_{j_2}\right)^2, \left(v_{j_2} - \tilde{v}^{(2)}_{j_2}\right)^2, \ldots, \left(v_{j_2} - \tilde{v}^{(q)}_{j_2}\right)^2\right\},$$

…………………………

$$j_{(k_\tau^*)} = n \text{ і } \varepsilon_{k_\tau^*} = \min\left\{\left(v_n - \tilde{v}^{(1)}_n\right)^2, \left(v_n - \tilde{v}^{(2)}_n\right)^2, \ldots, \left(v_n - \tilde{v}^{(q)}_n\right)^2\right\}.$$

Визначимо множини $I_1, I_2, \ldots, I_{k_\tau^*}$ в такому вигляді:

$$I_1 = \left\{i \in \{1, 2, \ldots, q\} | \varepsilon_1 = \left(v_{j_1} - \tilde{v}^{(i)}_{j_1}\right)^2\right\},$$

$$I_2 = \left\{i \in \{1, 2, \ldots, q\} | \varepsilon_2 = \left(v_{j_2} - \tilde{v}^{(i)}_{j_2}\right)^2\right\},$$

………………………………

$$I_{k_\tau^*} = \left\{i \in \{1,2,\ldots,q\} | \varepsilon_{k_\tau^*} = \left(v_n - \tilde{v}^{(i)}_n\right)^2\right\}$$

і побудуємо наступну таблицю

Табл. 2. Параметри прогнозуючої схеми

Моделі прогнозування	j_1	j_2	$\ldots$	$j_{k_\tau^*}$	Результуючий стовпчик
M_1	a_{11}	a_{12}	$\ldots$	$a_{1k_\tau^*}$	$S_1(\beta)$
M_2	a_{21}	a_{22}	$\ldots$	$a_{2k_\tau^*}$	$S_2(\beta)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_q	a_{q1}	a_{q2}	$\ldots$	$a_{qk_\tau^*}$	$S_q(\beta)$

де $a_{ps} = \begin{cases} \beta^{k_\tau^* - s}, \text{якщо } s \in I_s \\ 0, \quad \text{якщо } s \notin I_s \end{cases}$, $s_p(\beta) = \sum_{j=1}^{k_\tau^*} a_{pj}$,

$$0 < \beta < 1, \left(p = \overline{1,q};\ s = \overline{1,k_\tau^*}\right).$$

За допомогою $s_p(\beta)$ і $S(\beta) = \sum_{p=1}^{q} S_p(\beta)$ визначимо вагові коефіцієнти прогнозуючих моделей M_p $(p \le q)$ з якими вони входять у таку прогнозуючу схему

$$\tilde{v}_{n+\tau} = \frac{S_1(\beta)}{S(\beta)}\tilde{v}^{(1)}_{n+\tau} + \frac{S_2(\beta)}{S(\beta)}\tilde{v}^{(2)}_{n+\tau} + \ldots + \frac{S_q(\beta)}{S(\beta)}\tilde{v}^{(q)}_{n+\tau}. \qquad (7)$$

Перед тим, як використати прогнозуючу схем (7), ще треба провести навчання цієї схеми відносно β. Для цього будуємо функціонал

$$L(\beta) = \sum_{i=1}^{k_\tau^*}\left(v_{ji} - \frac{S_1(\beta)}{S(\beta)}\tilde{v}^{(1)}_{ji} - \frac{S_2(\beta)}{S(\beta)}\tilde{v}^{(2)}_{ji} - \ldots - \frac{S_q(\beta)}{S(\beta)}\tilde{v}^{(q)}_{ji}\right)^2,$$

$$\left(j_i = n - k_\tau^* + i\right)$$

і мінімізуємо його шляхом варіюванням значення β. Для цього інтервал (0,1] розбиваємо на m рівновеликих інтервалів і знаходимо значення $L(\beta_i)$ в точках $\beta_i = i/m,\ i = \overline{1,m}$. Очевидно, що через параметр m задаємо точність знаходження мінімуму функціоналу $L(\beta)$. Нехай $\beta_m^* = \min L(\beta_i)$. Тоді прогноз часового ряду проводимо за схемою (7), поклавши замість β величину β_m^*.

2. ***Метод навчання комбінованої моделі прогнозування часових рядів.*** Точність моделі (7) можна покращити внаслідок її навчання. Навчання моделі (7) проводимо за таким алгоритмом:

Крок 1. Задаємо m точок $\beta_r = r/m, r = \overline{1,m}$ в інтервалі (0,1] і для кожного фіксованого $\beta = \beta_r$ з ненульових елементів результуючого стовпчика табл. 2 побудуємо упорядковану множину

$$U_r = \left(S_{r_1}(\beta_r) \ge S_{r_2}(\beta_r) \ge \ldots \ge S_{r_{q_r}}(\beta_r)\right).$$

Крок 2. Визначимо множину найвпливовіших моделей U_r^* відносно $\beta = \beta_r$. На початку приймемо, що $U_r^* = \{M_{r_1}\}$. Рівність (7) відносно одної моделі M_{r_1} запишеться так: $\tilde{v}_{n+\tau} = \tilde{v}^{(r_1)}_{n+\tau}$.

Знаходимо значення функціоналу:

$$H_{r_1}(\beta_r) = \sum_{i=1}^{k_\tau^*}\left(v_{ji} - \tilde{v}^{(r_1)}_{ji}\right)^2.$$

Задаємо правило внесення моделі M_{r_2} у комбіновану модель. Розглянемо модель

$$\tilde{v}_{n+\tau} = \frac{S_{r_1}(\beta_r)}{S(\beta_r)}\tilde{v}^{(r_1)}_{n+\tau} + \frac{S_{r_2}(\beta_r)}{S(\beta_r)}\tilde{v}^{(r_2)}_{n+\tau}, \qquad (8)$$

де $S(\beta_r) = S_{r_1}(\beta_r) + S_{r_2}(\beta_r)$.

Знаходимо значення функціоналу $H_2(\beta_r)$ відносно моделі (8), а саме:

$$H_{r_2}(\beta_r) = \sum_{i=1}^{k_\tau^*}\left(v_{ji} - \frac{S_{r_1}(\beta_r)}{S(\beta_r)}\tilde{v}^{(r_1)}_{ji} - \frac{S_{r_2}(\beta_r)}{S(\beta_r)}\tilde{v}^{(r_2)}_{ji}\right)^2.$$

Якщо $H_2(\beta_r) < H_1(\beta_r)$, тоді модель M_{r_2} вносимо у множину U_r^*, тобто $U_r^* = \{M_{r_1}\} \cup \{M_{r_2}\}$. Процес побудови множини U_r^* для кожного фіксованого r продовжуємо до виконання умов внесення моделей із U_r. Після побудови множин U_r^* $(r = 1,2,\ldots,m)$ переходимо до кроку 3.

Крок 3. Нехай $U_r^* = \{M_{r_1}, M_{r_2}, \ldots, M_{r_n}\}$ і $H_{r_{h_r}^*}(\beta_{r^*}) = \min\{H_{r_{h_r}}(\beta_r) | r = \overline{1,m}\}$. Тоді комбінована прогнозуюча модель відносно множини моделей $U_{r^*}^* = \{M_{r_1^*}, M_{r_2^*}, \ldots, M_{r_{h_r}^*}\}$ має такий вигляд:

$$\tilde{v}_{n+\tau} = \frac{S_{r_1^*}(\beta_{r^*})}{S(\beta_{r^*})}\tilde{v}_{n+\tau}^{(r_1^*)} + \ldots + \frac{S_{r_{h_r}^*}(\beta_{r^*})}{S(\beta_{r^*})}\tilde{v}_{n+\tau}^{(r_{h_r}^*)}, \quad (9)$$

де $S(\beta_{r^*}) = S_{r_1^*}(\beta_{r^*}) + S_{r_2^*}(\beta_{r^*}) + \ldots + S_{r_{h_r}^*}(\beta_{r^*})$.

Важливим етапом прогнозування є верифікація прогнозів, тобто оцінювання їх точності та їх обґрунтованості. На етапі верифікації використовують сукупність критеріїв, які дають можливість оцінити якість прогнозу.

Ефективність прогнозуючої моделі (9) показано на даних пасажирських перевезень залізничним, річковим і автомобільним транспортами в Україні за період 1980-2013 [14].

Для оцінювання якості комбінованої моделі прогнозування застосовували критерій MRE (англ. *Mean Relative Error*) – середня відносна похибка. MRE визначається за формулою

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{v_t - \tilde{v}_t}{v_t}\right|,$$

де v_t – значення часового ряду у момент часу *t*; $\tilde{v}_t$ – прогнозне значення v_t.

Середню відносну похибку (MRE) можна використовувати для порівняння двох (або більше) різних прогнозів одного й того ж часового ряду: кращим вважається той прогноз, у якого значення MRE є меншим.

Табл. 3. Похибки прогнозу обсягу пасажирських перевезень за критерієм MRE

Методи прогнозування	Види пасажирських перевезень		
	Залізничний	Річковий	Автомобільний
Крок прогнозу $\tau = 1$			
Метод авторегресії	0,0041	0,0148	0,0115
Метод найменших квадратів з вагами	0,015	0,7975	0,1680
Лінійна модель Брауна	0,0358	0,0917	0,1478
Квадратична модель Брауна	0,0159	0,5516	0,086
Комбінована модель	0,0039	0,0148	0,0115
Крок прогнозу $\tau = 5$			
Метод авторегресії	0,0045	0,0111	0,0233
Метод найменших квадратів з вагами	0,0048	0,0683	0,0595
Лінійна модель Брауна	0,0585	0,0757	0,1482
Квадратична модель Брауна	0,0317	0,2295	0,0797
Комбінована модель	0,0031	0,0108	0,0225

Обговорення результатів дослідження. За критерієм середньої відносної похибки оцінимо якість прогнозу комбінованої прогнозуючої моделі шляхом порівняння її результатів з результатами прогнозуючих моделей з базової множини: M_1 – авторегресії, M_2 – метод найменших квадратів з вагами, M_3 – метод Брауна 1-го порядку, M_4 – метод Брауна 2-го порядку на підставі якої вона будується.

На підставі даних табл. 3 можна зробити висновок, що за середньою відносною похибкою точність прогнозу комбінованої моделі не уступає точності жодній із базових моделей прогнозування і здебільшого є найточнішою. На підставі цього можемо зробити висновок, що для прогнозування пасажирських перевезень залізничним, річковим та автомобільним транспортами в Україні доцільно використовувати комбіновану модель прогнозування.

Висновки

В умовах ринкової економіки серед, усіх чинників суспільного виробництва, людський чинник відноситься до найважливіших та найкреативніших. З огляду на це перевезення людей, зайнятих у процесі виробництва, надання послуг та реалізації продукції, відноситься до одної з найбільш актуальною функцією сьогоднішньої транспортної системи України. Забезпечуючи основну частину трудових поїздок населення та реалізацію права громадян на вільне переміщення, залізничний, річковий і автомобільний пасажирські транспорти безпосередньо впливають на ефективність економіки країни загалом. Якщо проаналізувати вплив пасажирського транспорту на економіки України, то варто відмітити, що пасажирський транспорт формує попит на продукцію таких галузей, як машинобудування, металургія, нафтовидобуток, нафтопереробка, електронна та інші, впливаючи на їхній економічний стан і динаміку.

У сучасних економічних умовах пасажирський транспорт перетворюється з інфраструктурної галузі у своєрідну провідну силу розвитку економіки, спроможний підвищити якість життя людей та активізувати господарську діяльність багатьох інших її галузей. З вищесказаного безпосередньо випливає, що розробка якісних і точних моделей прогнозування має важливе значення при побудові обґрунтованих планів розвитку промислових, залізничних, річкових і автотранспортних підприємств.

Розроблена комбінована модель була апробовано на реальних економічних показниках і показала високу точність прогнозування. Отже, результати прогнозу комбінованої моделі можна ефективно використовувати при розробці перспективних планів розвитку цілих галузей економіки України.

References

[1] Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews, 29*(5–6), 594–621. https://doi.org/10.1080/07474938.2010.481556

[2] Boxing, J., & Jenkins, G. (1974). Time series analysis. *Forecast and management*. Vol. 1. Moscow: Peace, 406. [In Russian].

[3] Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: *Statistical Mechanics and its Applications, 519*, 127–139. https://doi.org/10.1016/j.physa.2018.11.061

[4] Dolgikh, S., & Mulesa, O. (2021). Covid-19 epidemiological factor analysis: Identifying principal factors with machine. *CEUR Workshop Proceedings, 2833*, 114–123. https://doi.org/10.1101/2020.06.01.20119560

[5] Geche, F., Batyuk, A., Mulesa, O., & Vashkeba, M. (2015). Development of effective time series forecasting model. International *Journal of Advanced Research in Computer Engineering & Technology, 4*(12), 4377–4386.

[6] Geche, F., Mulesa, O., Batyuk, A., & Voloshchuk V. (2020). The Combined Time Series Forecasting Model, *IEEE Firs International Conference on Data Stream Mining & Processing* (DSMP), August 21–25, Lviv, Ukraine. 272–275. https://doi.org/10.1109/DSMP47368.2020.9204311

[7] Ivanov, V. V. (1999). *Time series analysis and forecasting of economic indicators*. Kharkiv: KhNU, 230 p. [In Russian].

[8] Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Procedia Computer Science, 48*, 173–179. https://doi.org/10.1016/j.procs.2015.04.167

[9] Kukharev, V. I., Sally, V. I., & Erpert, A. M. (1991). *Economic and mathematical methods and models in planning and management*. Kyiv: High school, 302. [In Russian].

[10] Litranovich, R. M. (2011). Construction and research of a mathematical model based on sources of experimental data by regression analysis. Rivne: IUEH, 140. [In Ukrainian].

[11] Mulesa, O. Yu., & Snytyuk, V. Ye. (2020). Development of an evolutionary method for time series forecasting. *Automation of technological and business processes, 12*(3), 4–9. [In Ukrainian]. https://doi.org/10.15673/atbp.v12i3.1854

[12] Shmueli, G., & Lichtendahl Jr, K. C. (2016). Practical time series forecasting with r: A hands-on guide. *Axelrod Schnall Publishers*.

[13] Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting, 36*(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.03.017

[14] Transport and Communications of Ukraine (2013). *State Statistics Service*. Statistical collection, 552 p. [In Ukrainian].

[15] Xu, W., Peng, H., Zeng, X., Zhou, F., Tian, X., & Peng, X. (2019). A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. *Applied Intelligence, 49*(8), 3002–3015. https://doi.org/10.1007/s10489-019-01426-3

F. E. Geche[1], O. Yu. Mulesa[1], A. Ye. Batyuk[2], V. Yu. Smolanka[1]

[1] Uzhhorod National University, Uzhhorod, Ukraine
[2] Lviv Polytechnic National University, Lviv, Ukraine

LEARNING A COMBINED MODEL OF TIME SERIES FORECASTING

The method of construction of the combined model of forecast ing of time series based on basic models of forecasting is developed in the work. The set of basic models is dynamic, ie new prediction models can be included in this set. Models also can be deleted depending on the properties of the time series. For the synthesis of a combined model of forecasting time series with a given forecast step, the optimal step of prehistory is determined at the beginning. Next the functional is constructed. The optimal prehistory step is determined using the autoregression method for a fixed forecast step. It determines the period of time at which the accuracy of models from the base set is analyzed. For each basic model during the process of the construction of the combined model is determined by the weighting factor with which it will be included in the combined model. The weights of the basic models are determined based on their forecasting accuracy for the time period determined by the prehistory step. The weights reflect the degree of influence of the base models on the accuracy of the combined model forecasting. After construction of the combined model, its training is carried out and those basic models which will be included in the final combined model of forecasting are defined. The rule of inclusion of basic models in the combined model is established. While including basic models in the combined forecasting model, their weights are taken into account, which depends on the same parameter. The optimal value of the parameter is determined by minimizing the given functional, which sets the standard deviation between the actual and predicted values of the time series. Weights with optimal parameters are ranked in decreasing order and are used to include basic models in the combined model.

As a result of this approach, as predicted values for the real time series show, it was possible to significantly improve the forecasting accuracy of the combined model in many cases. The developed method of training provides the flexibility of the combined model and its application to a wide class of time series. The results obtained in this work contribute to solving the problem of choosing the most effective basic models by synthesizing them into one combined model.

Keywords: time series; functional; forecasting model; model training; forecast step; weight coefficient.

Інформація про авторів:

Гече Федір Елемирович, д-р техн. наук, професор, в.о. завідувача кафедри кібернетики і прикладної математики.
Email: fedir.geche@uzhnu.edu.ua; https://orcid.org/0000-0002-4757-9828; ResearcherID: G-3504-2019

Мулеса Оксана Юріївна, канд. техн. наук, доцент, кафедра кібернетики та прикладної математики.
Email: oksana.mulesa@uzhnu.edu.ua; https://orcid.org/0000-0002-6117-5846; ResearcherID: F-6221-2019

Батюк Анатолій Євгенович, канд. техн. наук, доцент, кафедра автоматизованих систем управління. **Email:** abatyuk@gmail.com; https://orcid.org/0000-0001-7650-7383; ResearcherID: S-4988-2017

Смоланка Вероніка Юріївна, провідний інженер, програміст ЦІТ УжНУ. **Email:** veronika.smolanka@uzhnu.edu.ua; https://orcid.org/0000-0002-8380-1967; ResearcherID: S-4422-7750

Цитування за ДСТУ: Гече Ф. Е., Мулеса О. Ю., Батюк А. Є., Смоланка В. Ю. Навчання комбінованої моделі прогнозування часових рядів. Український журнал інформаційних технологій. 2021, т. 3, № 1. С. 44–48.

Citation APA: Geche, F. E., Mulesa, O. Yu., Batyuk, A. Ye., & Smolanka, V. Yu. (2021). Learning a combined model of time series forecasting. *Ukrainian Journal of Information Technology*, *3*(1), 44–48. https://doi.org/10.23939/ujit2021.03.044