DOI: https://doi.org/10.20535/kpisn.2021.2.236936
УДК 004.942:519.216.3

А.О. Белас*, П.І. Бідюк
КПІ ім. Ігоря Сікорського, Київ, Україна
*corresponding author: andrii.belas@gmail.com

ВИБІР КРИТЕРІЮ ЯКОСТІ ДЛЯ ОЦІНЮВАННЯ ПРОГНОЗІВ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

Проблематика. Прогнозування нелінійних нестаціонарних процесів (ННП), поданих у формі часових рядів, є актуальним, оскільки такі ряди можуть описувати процеси в технічних й економічних системах. Для вибору найкращої математичної моделі використовують різні метрики оцінювання якості прогнозів, як-от: R^2, RMSE, MAE, MAPE. Однак оптимізація моделі за одним критерієм погіршує її відносно іншого. Тому важливо розуміти, яку метрику слід використовувати для оптимізації та оцінки якості прогнозу в поставленій задачі.
Мета дослідження. Розробити та проаналізувати критеріальну базу для оцінювання прогнозів ННП, а також підхід до вибору критерію якості з урахуванням особливостей поставленої задачі прогнозування.
Методика реалізації. Виконання порівняльного аналізу основних метрик для задачі регресії, а саме їхнього теоретичного та практичного змісту, переваг і недоліків у різних випадках.
Результати дослідження. Сформовано критеріальну базу для оцінювання прогнозів ННП, а також підхід до вибору критерію якості з урахуванням особливостей поставленої задачі прогнозування. Для мінімізації абсолютної похибки проаналізовано та рекомендовано використовувати метрики RMSE (MSE, R^2) і MAE залежно від необхідності роботи з викидами. Для розв'язання задач мінімізації відносної помилки запропоновано використовувати метрику RMSLE.
Висновки. Показано важливість вибору метрики для оптимізації та оцінювання якості прогнозу в поставленій задачі. Отримані критеріальну базу та підхід можна використовувати в подальших дослідженнях як для розв'язання практичних задач моделювання та прогнозування ННП, так і для розробки нових методів або загальної методики розв'язання цих задач.
Ключові слова: математична модель; прогнозування; критерій якості прогнозів; часовий ряд; метрика; нелінійний нестаціонарний процес.

Вступ

Задача прогнозування нелінійних нестаціонарних процесів (ННП) на основі експериментальних даних, представлених у формі часових рядів, є актуальною і має численні застосування в технічних, соціально-економічних, екологічних і фінансових системах [1].

У роботі [2] сформульовано й описано загальну методику прогнозування ННП на основі математичних моделей. Одним із важливих етапів методики є оцінювання якості отриманих прогнозів за допомогою певних критеріїв і добір кращої з оцінених моделей-кандидаток. Лише після цього математичну модель застосують для розв'язання основних завдань – прогнозування, керування та поглибленого дослідження процесу. Розглянуто найпоширеніші статистичні критерії якості прогнозу: середньоквадратичну похибку (RMSE), середню абсолютну похибку (MAE), середню абсолютну похибку у відсотках (MAPE), середньоквадратичну логарифмічну похибку (RMSLE).

Можна очікувати, що найкраща математична модель зможе отримати найкращі результати за всіма доступними метриками, проте досить часто це не так. Зазвичай на початку оптимізації структури та параметрів моделі більшість критеріїв поліпшуються, проте з часом досягають точки, коли покращення стану однієї метрики може призвести до погіршення іншої. У таких випадках виникає проблема добору пріоритетного критерію оцінювання якості прогнозів.

Розглянемо поширені метрики для задачі прогнозування, їхні переваги та недоліки в різних випадках; сформуємо рекомендації щодо вибору критерію якості відповідно до особливостей певної задачі прогнозування на основі реальних даних із прогнозування продажів.

Для прикладів використано вибірку з фактичних даних із продажів для 45-ти магазинів "Walmart", розташованих у різних регіонах [3]. Розв'язано задачу прогнозування щотижневих продажів у різних магазинах і відділах роздрібної торгівлі. Обсяг вибірки – 140 тижнів, з яких для тестування й оцінювання якості прогнозів використано останні 26.

Як загальну методику обрано запропоновану в [2]. Основними методами аналізу даних будуть ті, що розглянуті в роботі [4].

Постановка задачі

Розробити критеріальну базу оцінювання якості прогнозів ННП, а також підхід до вибору критерію якості з урахуванням особливостей поставленої задачі прогнозування.

Метрики оцінювання якості прогнозів

Важливим моментом прогнозування є об'єктивне визначення якості прогнозу. Зазвичай дослідників цікавить, як добре алгоритм прогнозування працює з даними, до яких не застосовувався раніше. Це показує, як модель поводитиметься з даними на практиці, та є метою машинного навчання – *узагальненням* його алгоритмів. Точність прогнозів можна визначити лише з урахуванням того, наскільки якісно модель працює з новими даними, які не використовували під час її навчання, тому вибірку даних доцільно розділити на навчальну та тестову. На навчальній вибірці виконують навчання – оцінювання структури та параметрів методу прогнозування – та реалізують "історичний" прогноз, який дає змогу встановити якість однокрокового прогнозу на певній ділянці ряду.

Прогноз на тестовій вибірці використовують для оцінювання його точності та вибору найкращої моделі, що має властивості узагальнення. Оскільки тестові дані не використовують для навчання моделі, вони повинні показати, як вдало модель прогнозує нові дані. Так запобігають ефекту "перенавчання". У різних емпіричних дослідженнях рекомендують залишати для перевірки 5–40 % значень ряду даних. Хоча під час аналізу коротких рядів доцільно використовувати значно більший відсоток ряду для оцінювання структури та параметрів моделі. Також це значення залежить від тривалості вибірки та того, настільки далеко необхідно прогнозувати. В ідеалі тестова вибірка повинна відповідати величині необхідного максимального горизонту прогнозу. Після вибору математичної моделі прогнозують значення поза вибіркою даних. Підходи до поділу вибірки на навчальну та тестову детально розглянуто у джерелах [5–8]. У цій роботі будемо вважати, що критерії якості обчислюють для тестового набору даних, окремого від тих, на яких система навчалася. Метрики якості прогнозування відрізняються від "залишків прогнозів", адже залишки розраховують на навчальній вибірці, тоді як похибки оцінювання – на тестовій. До того ж залишки ґрунтуються на однокрокових прогнозах, а похибки прогнозів можуть містити багатокрокові прогнози.

Для вибору найкращої моделі під час розв'язання певної задачі використовують різні метрики оцінювання якості прогнозів.

Метрики RMSE, MSE, R^2

Часто сума похибок прогнозів дорівнює нулю, оскільки похибки мають різні знаки, а тому для них необхідно використовувати інші міри. Наприклад, застосовувати квадрати похибок. Отримана формула за змістом схожа на стандартне відхилення, але прогнозних значень відносно реальних.

Середньоквадратична похибка моделі (англійською RMSE – Root Mean Squared Error) моделі є дуже популярною метрикою й обчислюється так:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{k}=1}^{N}[\hat{y}(k) - y(k)]^2}{N}},$$

де (k) - – прогнозоване за моделлю значення; $y(k)$ – реальне вимірювання; N – довжина вибірки.

RMSE може набувати значень $[0,+\infty)$. Чим вона менше, тим кращим вважають прогноз. Усі описані нижче властивості стосуються також спорідненої середньоквадратичної помилки MSE (Mean Squared Error):

$$\mathrm{MSE} = \frac{1}{N}\left[y(k) - \hat{y}(k)\right]^2, \ \mathrm{RME} = \sqrt{\mathrm{MSE}}.$$

$$\mathrm{MSE}(a) > \mathrm{MSE}(b) \Leftrightarrow \mathrm{RMSE}(a) > \mathrm{RMSE}(b).$$

MSE є евклідовою відстанню прогнозів до реальних значень. Також цю метрику можна отримати за методом максимальної вірогідності. Для цього розглянемо прогнозну модель як

$$y = a_w(x) + \varepsilon, \ \varepsilon\pi \sim N(0, \sigma^2),$$

де y – фактичні виміри; $a_w(x)$ – модель зі змінними x і параметрами w; ε – нормально розподілена випадкова величина (шум) із заданою фіксованою дисперсією σ^2.

Для оцінювання параметрів w моделі $a_w(x)$ випишемо функцію вірогідності моделі:

$$p(y \mid x, w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - a_w(x))^2}{2\sigma^2}\right].$$

Застосуємо метод максимальної вірогідності

$$\log L(\mathrm{w}) = \log \prod_{i=1}^{m} p(y_i \mid x_i, w_i) =$$

$$= \sum_{\mathrm{i}=1}^{m} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - a_w(x_i))^2}{2\sigma^2} \right] \to \max,$$

де m — розмір вибірки. Звідси отримуємо задачу мінімізації MSE:

$$\frac{1}{2\sigma^2} \sum_{\mathrm{i}=1}^{m} (y_i - a_w(x_i))^2 \to \min,$$

де MSE мінімізує як зміщення точкової оцінки, так і дисперсію.

Якщо втрати під час прогнозування, пов’язані з завищенням фактичного майбутнього значення, врівноважуються заниженням, то ідеальний прогноз повинен бути незміщеним. У такому разі MSE має прямувати до нуля.

Зміщення точкової оцінки визначається як

$$Bias(\hat{\theta}_m) = E(\hat{\theta}_m) - \theta,$$

де θ — випадкова величина.

Оскільки MSE вимірює очікувану розбіжність (у контексті квадратів помилок), то маємо:

$$\mathrm{MSE} = E[(\hat{\theta}_m - \theta)^2] = Bias(\hat{\theta}_m)^2 + Var(\hat{\theta}_m).$$

Значна кількість способів отримання критерію якості MSE пояснює його популярність.

Основні властивості MSE:

1. Штраф стосовно великих похибок збільшується, а для малих зменшується. Менші похибки (наприклад, менші за 1) матимуть ще менший внесок у загальну похибку після піднесення до квадрата, тоді як більші похибки матимуть набагато більшу вагу після піднесення до квадрата.

2. Чутливість до викидів. Аналогічно до попереднього пункту наявність викидів (аномально великих значень даних) призведе до вагомих значень метрики, що також вплине на поведінку алгоритму оптимізації.

3. Легка оптимізація. Структура функції, завдяки використанню квадрата, легко диференційована, тому різні алгоритми (наприклад, градієнтний спуск) зможуть із нею працювати. Хоча метрики MSE та RMSE споріднені, їхні градієнти відрізнятимуться:

$$\frac{\partial \mathrm{MSE}}{\partial a} = \frac{2}{m} \sum_{i=1}^{m} (a(x_i) - y_i),$$

$$\frac{\partial \mathrm{RMSE}}{\partial a} = \frac{1}{m\mathrm{RMSE}} \sum_{i=1}^{m} (a(x_i) - y_i),$$

де $a(x)$ — прогнозна модель зі змінними x; m — розмір вибірки. Тобто градієнти відрізняються в 2/RMSE рази.

4. Оптимальне розв’язання задачі мінімізації RMSE у класі констант — це середнє арифметичне ряду, як показано в [9].

Ці властивості також стосуються метрики R^2, оскільки її формулу можна виразити через MSE:

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE}(\bar{y})},$$

де $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ — модель, яка прогнозує все середнім значенням навчальної вибірки.

Тобто метрика R^2 показує, наскільки наша модель (за критерієм MSE) краща за наївну модель, що прогнозує все середнім значенням. Якщо наша модель не припускається помилок, то R^2 = 1; якщо наша модель така ж, як і наївна модель, то R^2 = 0. Однак вона може бути й гірша, тому діапазон можливих значень для метрики R^2: $(-\infty,1]$, і вважається, що чим ця метрика більша, тим краще.

Отже, RMSE, MSE, R^2 — дуже популярні метрики. Їх доцільно застосовувати у випадках, коли не можемо дозволити собі велику похибку в жодному зі спостережень. Інакше кажучи, нам може бути зручно мати трохи більшу похибку на багатьох зразках, якщо ніколи не отримаємо занадто велику похибку на жодному з них. Наприклад, якщо з погляду постановки задачі похибка прогнозу на 200 понад удвічі гірша, ніж похибка на 100, то слід використовувати RMSE чи MSE.

Метрика MAE

Замість квадратів можна використовувати модулі похибок, що також розв’яже проблему різних знаків у значеннях похибки (залежно від перепрогнозу чи недопрогнозу). Це теж актуальний підхід, який часто застосовують із MSE. Отримана метрика називатиметься середньою абсолютною похибкою, або MAE (Mean Absolute Error) й обчислюватиметься за формулою

$$\mathrm{MAE} = \frac{\sum_{k=1}^{N} |\hat{y}(k) - y(k)|}{N},$$

де (k) - — прогнозоване за моделлю значення; $y(k)$ — реальне вимірювання; N — довжина вибірки.

MAE може набувати значень $[0,+\infty)$ та є дуже інтерпретабельною, оскільки вимірюється в тих самих одиницях, що й вихідний ряд. Чим меншою є MAE, тим кращим вважають прогноз. Зазначимо, що математично MAE є мангеттенською метрикою, введеною Германом Мінковським [10]. Також можна отримати теоретичне підґрунтя для використання цієї метрики як функції похибки за допомогою методу максимальної вірогідності. Нехай модель будують як і раніше, але похибки (шум) будемо вважати розподіленими за Лапласом:

$$y = a_w(x) + \varepsilon, \varepsilon \sim laplace(0, \alpha).$$

Для оцінювання параметрів w моделі $a_w(x)$ випишемо вірогідність моделі

$$p(y \mid x, w) = \frac{\alpha}{2} \exp[-\alpha \mid y - a_w(x) \mid].$$

Застосування методу максимальної вірогідності відповідає максимізації

$$\log L(w) = \log \prod_{i=1}^{m} p(y_i \mid x_i, w) =$$

$$= \sum_{i=1}^{m} \log\left[\frac{\alpha}{2} - \alpha \mid y_i - a_w(x_i) \mid\right] \to \max$$

або (відкидаючи константи)

$$\alpha \sum_{i=1}^{m} \mid y_i - a_w(x_i) \mid \to \min.$$

Тож максимізація функції вірогідності еквівалентна мінімізації MAE.

Основними властивостями MAE є:

1. Похибки зважуються однаково, похибка на 2 гірша за похибку на 1 удвічі.

2. Значно стійкіша до викидів, ніж MSE. Якщо ми впевнені, що в даних це саме викиди (помилкової природи), а не певні великі значення, цікаві з погляду дослідження, — варто використовувати MAE, а не MSE.

3. Складність оптимізації. MAE не диференційована в нулі (коли прогнози рівні фактичним значенням), і, залежно від розподілу ряду, це може зробити різні наближення для MAE кращими за інші.

4. Оптимальне розв'язання задачі мінімізації MAE в класі констант — це медіана вихідного ряду, як показано в [9]. Для задач "переривчастого попиту", коли в ряді багато нульових значень, небезпечно використовувати MAE, оскільки мінімізація за нею прямуватиме до того, щоб вважати нулем усі прогнози (бо вони будуть прямувати до медіани).

Отже, MAE є також популярною метрикою. Її варто використовувати в протилежних для RMSE випадках. Тобто коли не хочемо підкреслювати великі похибки, а навпаки — вважаємо похибки однаково важливими за своїми обсягами. MAE більш стійка до викидів, ніж RMSE.

Метрика MAPE

Розглянемо впровадження відносної метрики на прикладі задачі прогнозування курсу акцій на біржі. У разі, коли реальний курс був \$ 1, а ми спрогнозували \$ 2, з погляду задачі дуже помилилися, а якщо реальний курс був \$ 100, а спрогнозували \$ 101, то насправді майже вгадали. Однак MAE всюди буде однакова. Тому було запропоновано розглянути різні відносні чи відсоткові метрики, незалежні від одиниці вимірювання ряду. Оскільки така міра характеризує відносну якість прогнозу, то її використовують переважно для порівняння точності прогнозів різнорідних об'єктів (процесів) прогнозування. Однак вона є також зручною для порівняльного аналізу якості прогнозування того самого процесу різними методами, оскільки відносна міра є чіткою та зрозумілою для дослідника та користувача [5, 6].

Найпопулярніша з таких метрик, середня абсолютна похибка у відсотках (Mean Absolute Percentage Error), обчислюється за формулою

$$\text{MAPE} = \frac{100\ \%}{N} \sum_{k=1}^{N} \left| \frac{\hat{y}(k) - y(k)}{y(k)} \right|,$$

де (k) - — прогнозоване за моделлю значення; $y(k)$ — реальне вимірювання; N — довжина вибірки. MAPE може набувати значень $[0,+\infty)$; чим менше значення метрики, тим кращий прогноз.

Основні властивості MAPE:

1. Похибки зважуються однаково, похибка в 20 % вдвічі гірша за похибку в 10 %.

2. Метрика стійка до викидів у даних.

3. Складність оптимізації.

4. Оптимальне розв'язання задачі мінімізації MAE в класі констант — це зважена медіана вихідного ряду, як показано в [9].

Проте ця метрика має значний недолік: якщо реальні значення будуть нульовими чи доволі близькими до нуля, то значення MAPE прямуватиме до нескінченності. Тоді, щоб обчислити цей критерій, нульові значення необхідно пропускати з відповідним коригуванням значення N [5]. Можливо, цей підхід дає змогу наближено проаналізувати якість прогнозування, проте ми не враховуємо нульові значення під час оцінювання якості, а прогнозування саме нульових значень може бути суттєвим, якщо навіть не принциповим. Тому метрику не можна застосувати за наявності нульових або близьких до нуля значень. Також за великого діапазону значень ряду, наявності "стрибків" угору чи вниз у цій метриці можна отримувати дуже великі значення, що робить її незастосовною. Наприклад, прогнозований дохід на завтра – \$ 1000000, а фактичний дохід став \$ 0,01. У такому разі MAPE буде 9999999900 %, і такі випадки – часте явище на практиці.

Негативні похибки штрафуються сильніше за позитивні [11], тому було запропоновано метрику SMAPE [12]:

$$\mathrm{SMAPE} = \frac{1}{N}\sum_{k=1}^{N}\frac{|\,\hat{y}(k) - y(k)\,|}{(|\,y(k)\,| - |\,\hat{y}(k)\,|)\,/\,2}100,$$

де (k) - – прогнозоване за моделлю значення; $y(k)$ – фактичний вимір; N – довжина вибірки.

Проте вона також не розв'язує проблему нулів і, за [13], проблему симетричності. У роботі [14] показано, що MAPE і SMAPE не рекомендують використовувати, тож запропоновано метрику MASE.

$$\mathrm{MASE} = \frac{1}{N}\sum_{i=1}^{N}|q_i|,$$

$$q_i = \frac{y(i) - \hat{y}(i)}{\dfrac{1}{N-1}\displaystyle\sum_{k=2}^{N}|y(k) - y(k-1)|}.$$

MASE є дієвою та водночас важкою для використання й інтерпретації, тому її частіше застосовують для наукових досліджень, ніж практики. У роботах [15, 16] запропоновано різні складні метрики для вирішення цієї проблеми. Наприклад, SMAE [15] розв'язує проблему нулів, проте питання різних рівнів, особливо, якщо дані нестаціонарні, є невирішеним.

Отже, MAPE є цікавою для обрахунку відносної похибки та простої інтерпретації, проте її можна застосовувати, лише тоді, коли дані перебувають далеко від нуля та не мають великого розкиду значень. Деякі з цих проблем можуть бути розв'язані вказаними вище відносними метриками, що є досить складними для використання та інтерпретації.

Метрика RMSLE

Середньоквадратична логарифмічна похибка (Root Mean Squared Logarithmic Error) моделі пропонується як метрика для мінімізації відношення до реальних значень (як і MAPE).

RMSLE обчислюють за формулою

$$\mathrm{RMSLE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[\log(y(k)+1) - \log(\hat{y}(k)+1)]^2},$$

де (k) - – прогнозоване за моделлю значення; $y(k)$– фактичний вимір; N – довжина вибірки.

Додавання одиниці (чи будь-якої константи) дає змогу врахувати випадки, коли цільова змінна дорівнює нулю (дає можливість обчислити логарифм). RMSLE може набувати значень $[0,+\infty)$; чим менше значення, тим кращий прогноз. Теоретичне обґрунтування RMSLE не відрізняється від RMSE, оскільки є тим самим RMSE, але над логарифмованим рядом, тобто

$$\mathrm{RMSLE} = \mathrm{RMSE}(\log(y(k)+1),\ \ \log(\hat{y}(k)+1)).$$

Логарифмування ряду також має згладжувальний ефект, зводячи значення ряду приблизно до одного рівня, що може допомогти за наявності експоненційних трендів. Однак головна ідея полягає в тому, що логарифм частки буде різницею логарифмів, а саме:

$$\mathrm{RMSLE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\log\frac{y(k)+1}{\hat{y}(k)+1}\right)^2},$$

роблячи RMSLE такою, що мінімізує відносну похибку, зберігаючи всі переваги класичної RMSE.

Основні властивості RMSLE:

1. Штраф до великих помилок збільшується, однак завдяки логарифмуванню вплив великих значень зменшується, тому метрику можна вважати збалансованою.

2. Нечутливість до викидів завдяки логарифмуванню – перевага, крім випадків, коли ці викиди потрібно знайти.

3. Легка оптимізація завдяки тому, що можна покращити MSE над логарифмованим рядом.

4. Оптимальне розв'язання задачі мінімізації RMSLE в класі констант – це середнє арифметичне логарифмованого ряду аналогічно до RMSE.

Отже, RMSLE є ідеальним варіантом, коли необхідно мінімізувати відносну помилку. Метрика не має недоліків MAPE, дуже зручна для практичного використання, має класичне теоретичне обґрунтування. Під час використання RMSLE варто пам'ятати про її нечутливість до викидів.

Практична частина

У таблиці 1 наведено приклад перших п'яти записів вибірки для першого магазину (згідно з [3]).

***Таблиця* 1**. Перші п'ять записів вибірки даних продажів першого магазину "Walmart"

Date	Weekly_Sales
2010-02-05	24924,5
2010-02-12	46039,49
2010-02-19	41595,55
2010-02-26	19403,54
2010-03-05	21827,9

Проведемо експеримент на всій вибірці даних, оптимізуючи параметри моделі так, щоб мінімізувати RMSE; виконаємо прогноз на тестовій вибірці та обрахуємо різні метрики (табл. 2).

***Таблиця* 2**. Значення різних метрик для оптимізованої за RMSE моделі

Метрика	Значення
MAE	2076,3
MAPE	24,196
MSE	1,3387e+07
R^2	0,95816
RMSE	**3658,8**
RMSLE	3,56
SMAPE	17,048

Проведемо експеримент на всій вибірці даних, оптимізуючи параметри моделі так, щоб мінімізувати MAE; виконаємо прогноз на тестовій вибірці та обрахуємо різні метрики (табл. 3).

***Таблиця* 3**. Значення різних метрик для оптимізованої за MAE моделі

Метрика	Значення
MAE	**1883,8**
MAPE	17,182
MSE	1,3847e+07
R^2	0,95734
RMSE	3721,1
RMSLE	4,26
RMSPE	6875,9
SMAPE	13,851

MAE стала кращою (1883 < 2076), ніж тоді, коли ми оптимізували модель за RMSE, проте RMSE погіршилась (3721 > 3658). Отже, модель, краща за однією метрикою, не гарантує кращого результату за всіма іншими метриками — саме тому доволі критичним є розуміння всіх основних критеріїв якості, їх переваг і недоліків для добору правильного для поставленої задачі критерію.

Висновки

Аналіз вибраних даних показав: оптимізація моделі для одного критерію погіршує її відносно іншого. Тому важливо розуміти, яку метрику використовувати для оптимізації та оцінки якості прогнозу в конкретній поставленій задачі.

Для мінімізації абсолютної похибки варто використовувати метрики RMSE (MSE, R^2) і MAE, залежно від необхідності роботи з викидами.

Розглянуті критеріальну базу та підхід можна використовувати в подальших дослідженнях як для розв'язання практичних задач моделювання та прогнозування ННП, так і для розробки нових методів або загальної методики розв'язання цих задач.

References

[1] A.K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications.* London, England: Springer-Verlag London, 2005, 372 p. doi: 10.1007/1-84628-184-9

[2] O.M. Belas *et al.*, "General approach to forecasting nonlinear nonstationary processes using mathematical models on statistical data," *Special Telecommun. Syst. Inform. Security*, vol. 2, no. 4. pp. 26–31, 2018.

[3] *Walmart Recruiting – Store Sales Forecasting* [Online]. Avaliable: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data.

[4] O.M. Belas *et al.*, "Comparative analysis of autoregressive approaches and recurrent neural networks for modeling and forecasting of nonlinear nonstationary processes," *Inform. Technol. Security*, vol. 7, no. 1, pp. 91–99, 2019.

[5] P.I. Bidyuk *et al.*, *Time Series Analysis.* Kyiv, Ukraine: Polytechnica, 2010, 317 p.
[6] R.J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice.* Melbourne, Australia: OTexts, 2013.
[7] I. Goodfellow *et al.*, *Deep Learning.* New York: MIT Press, 2016, 800 p.
[8] C. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer-Verlag New York, 2006, 738 p.
[9] S. Kolassa, "Evaluating predictive count data distributions in retail sales forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 788–803, 2016. doi: 10.1016/j.ijforecast.2015.12.004
[10] E.F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry.* New York: Dover Publications, 1987, 96 p.
[11] J. S. Armstrong, *Long-range Forecasting: From Crystal Ball to Computer.* New York: John Wiley & Sons, 1978.
[12] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *Int. J. Forecasting*, vol. 9, no. 4, pp. 527–529, 1993. doi: 10.1016/0169-2070(93)90079-3
[13] P. Goodwin and R. Lawton, "On the asymmetry of the symmetric MAPE," *Int. J. Forecasting*, vol. 15, no. 4, pp. 405–408, 1999. doi: 10.1016/S0169-2070(99)00007-2
[14] R.J. Hyndman and A.B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006. doi: 10.1016/j.ijforecast.2006.03.001
[15] F. Petropoulos and N. Kourentzes, "Forecast combinations for intermittent demand," *J. Operat. Res. Soc.*, vol. 66, no. 6, pp. 914–924, 2015, doi: 10.1057/jors.2014.62
[16] A. Davydenko and R. Fildes, "Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts," *Int. J. Forecasting*, vol. 29, no. 3, pp. 510–522, 2013. doi: 10.1016/j.ijforecast.2012.09.002

А.О. Белас, П.И. Бидюк

ВЫБОР КРИТЕРИЯ КАЧЕСТВА ДЛЯ ОЦЕНИВАНИЯ ПРОГНОЗОВ НЕЛИНЕЙНЫХ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ

Проблематика. Прогнозирование нелинейных нестационарных процессов (НПП), представленных в форме временных рядов, является актуальным, поскольку такие ряды могут описывать процессы в технических и экономических системах. Для выбора лучшей математической модели используют различные метрики оценивания качества прогнозов, такие как: R^2, RMSE, MAE, MAPE. Однако оптимизация модели по одному критерию ухудшает ее относительно другого. Поэтому важно понимать, какую метрику следует использовать для оптимизации и оценивания качества прогноза в поставленной задаче.

Цель исследования. Разработать и проанализировать критериальную базы для оценивания прогнозов НПП, а также подход к выбору критерия качества с учетом поставленной задачи прогнозирования.

Методика реализации. Выполнение сравнительного анализа основных метрик для задачи регрессии, а именно их теоретического и практического смысла, преимуществ и недостатков в различных случаях.

Результаты исследования. Сформирована критериальная база для оценивания прогнозов ННП, а также подход к выбору критерия качества с учетом особенностей поставленной задачи прогнозирования. Для минимизации абсолютной погрешности проанализированы и рекомендованы к использованию метрики RMSE (MSE, R^2) и MAE в зависимости от необходимости работы с выбросами. Для решения задач минимизации относительной погрешности предложено использовать метрику RMSLE.

Выводы. Показана важность выбора метрики для оптимизации и оценивания качества прогноза в поставленной задаче. Полученные критериальную базу и подход можно использовать в дальнейших исследованиях как для решения практических задач моделирования и прогнозирования ННП, так и для разработки новых методов или общей методики решения таких задач.

Ключевые слова: математическая модель; прогнозирование; критерий качества прогнозов; временной ряд; метрика, нестационарный процесс.

A.O. Belas, P.I. Bidyuk

CHOOSING A QUALITY CRITERION FOR EVALUATING THE FORECAST OF NONLINEAR NON-STATIONARY PROCESSES

Background. The problem of forecasting nonlinear nonstationary processes presented in the form of time series is very relevant, since such series can describe dynamics of the processes in both technical and economic systems. To establish the best model, various metrics are used to assess the quality of forecasts, such as R^2, RMSE, MAE, MAPE. However, in many tasks, when optimizing the model according to the selected criterion, the model becomes worse in relation to another criterion. Therefore it is important to understand which metric must be used to optimize and assess the quality of the forecast in the given task.

Obective. The aim of the paper is to develop a criteria base for assessing forecasts of nonlinear nonstationary processes, as well as an approach to choosing a metric in accordance to the specificity of the set forecasting problem.

Methods. The paper presents a comparative analysis of the basic metrics for the regression problem, their theoretical and practical meaning, advantages and disadvantages in various cases. New approaches are proposed based on the results of the analysis.

Results. Based on the analysis of the selected data, it is shown that by optimizing the model according to the selected criterion, the model becomes worse in relation to another criterion. A criterion basis for assessing forecasts of nonlinear nonstationary processes has been formed, as well as an approach to the selection of a quality criterion in accordance with the specifics of the set forecasting problem. To minimize an absolute error, the RMSE (MSE, R^2) and MAE metrics are analysed and recommended, depending on the need to work with outliers. The RMSLE metric is proposed for solving the problems of minimizing the relative metric, for solving the shown problems of the MAPE metric for this class of problems.

Conclusions. The paper shows the importance of choosing a metric that must be used to optimize and assess the quality of the forecasts in the given task. The obtained criterion base and approach can be used in further research to solve practical problems in modelling and forecasting nonlinear nonstationary processes and to develop new methods or general method for solving such problems.

Keywords: mathematical model; forecasting; forecast quality criteria; time series; metrics; non-stationary process.

Рекомендована Радою
Інституту прикладного системного аналізу
КПІ ім. Ігоря Сікорського

Надійшла до редакції
11 квітня 2021 року

Прийнята до публікації
14 червня 2021 року