Метод побудови прогнозуючої моделі з динамічними параметрами

В.М.Синєглазов
кафедра авіаційних комп'ютерно-інтегрованих комплексів
Інститут інформаційно-діагностичних систем
Київ, Україна
svm@nau.edu.ua

О.І. Чумаченко
кафедра технічної кібернетики
НТУУ «КПІ»
Київ, Україна
lobach21@mail.ru

В.С. Горбатюк
кафедра технічної кібернетики
НТУУ «КПІ»
Київ, Україна
vladislav.horbatiuk@gmail.com

Method of building the forecasting model with dynamic parameters

V. Sinyeglazov
Department of Aviation Computer-Integrated Complexes
Institute of Information and Diagnostic Systems
Kyiv, Ukraine
svm@nau.edu.ua

O. Chumachenko
Department of Technical Cybernetics
NTUU "KPI"
Kyiv, Ukraine
lobach21@mail.ru

V. Gorbatiuk
Department of Technical Cybernetics
NTUU "KPI"
Kyiv, Ukraine
vladislav.horbatiuk@gmail.com

***Анотація*—Із задачею прогнозування пов'язані такі проблеми як непостійність впливу вхідних факторів на прогнозований процес, велика кількість придатних моделей та інші. У цій роботі запропоновано нову постановку задачі прогнозування та метод побудови прогнозуючої моделі, який динамічно вибирає набір моделей в залежності від заданих входів і враховує мінливий характер значущості / впливу факторів. Запропонований метод було порівняно з 2 іншими методами побудови прогнозуючих моделей на 11 різних наборах даних, і він мав найменшу середню середньоквадратичну помилку.**

***Abstract*—The forecasting problem is rightly considered to be one of the most important and interesting tasks. It is also one of the most difficult ones, since it is associated with such problems as the impact of volatility of input factors on the forecast process, a large number of suitable models and others. In this paper a method that tries to deal with the first two questions is suggested, as it dynamically selects a set of models based on specified inputs and takes into account the changing nature of factors significance / impact. The approach designed by the authors allows to move from the global to globally-local methods, allowing to receive more qualitative forecasting model (in terms of minimum error estimation of the forecasting model), applying it to the existing global practices. The described algorithm was tested on a set of samples that are publicly available on the Internet. The proposed method was compared with two other methods of forecasting models building for 11 different sets of data, and it had the lowest average mean square error.**

Ключові слова*—*алгоритм прогнозування, прогнозування часових рядів, штучні нейронні мережі, неоднорідна вибірка, динамічні параметри

Keywords*—*forecasting algorithm, time series forecasting, artificial neural networks, homogeneous sample, dynamic parameters

I. Вступ

Задачу прогнозування по праву можна вважати однією з найбільш важливих і цікавих проблем. Вона також є однією з найскладніших задач, так як при її вирішенні доводиться стикатися з такими труднощами:

• неможливо врахувати всі чинники, що впливають на прогнозований процес; більш того, їх вплив може змінюватися з плином часу - фактор, який не був важливий сьогодні, може грати критичну роль завтра;

• зазвичай існує нескінченне число можливих моделей, які добре відповідають наявним даним - необхідно вирішити, яку модель або множину моделей використовувати;

• часто буває важко (якщо не неможливо) знайти оптимальну складність моделі.

У даній роботі запропоновано метод, який намагається впоратися з першими двома питаннями: він динамічно вибирає набір моделей в залежності від заданих входів і враховує мінливий характер значущості / впливу факторів.

II. Постановка задачі

При постановці задачі прогнозування найчастіше робиться припущення, що характер прогнозованого процесу на спостережуваному періоді не змінюється, і, таким чином, завдання можна вирішити шляхом знаходження потрібної моделі і оцінки її параметрів, використовуючи всі наявні дані. Однак найчастіше поведінка прогнозованого процесу може кілька разів змінюватися протягом спостережуваного періоду, що робить таку постановку задачі некоректною - одна єдина модель не зможе описати кілька різних станів процесу. Розглянемо постановку задачі прогнозування, яка явно враховує проблему непостійного характеру прогнозованого процесу.

Нехай є:

а) множина прикладів $S = \{(\vec{x}_i, y_i) : \mathrm{i} = 1 \ldots \mathrm{n}\}$, де кожен приклад $(\vec{x}_i, y_i)$ складається з вектора значень вхідних факторів і відповідного вихідного значення;

б) модель f, яка залежить від деякого вектора параметрів $\vec{\theta} = [\theta_1, \ldots, \theta_m]^T$ і найбільш загальним способом описує прогнозований процес - мається на увазі, що в залежності від значень параметрів модель здатна описати всі можливі стани об'єкта, який прогнозується.

Необхідно знайти:

а) множину векторів параметрів $\Theta = \{\vec{\theta}_1, \ldots, \vec{\theta}_k\}$, таких, що вектор $\vec{\theta}_j$ є оптимальним згідно критерію L_j на деякій підмножині вихідних даних S_j: $\Theta = \left\{ \vec{\theta}_j : \vec{\theta}_j = \underset{\theta}{\operatorname{argmin}}(L_j[f, \vec{\theta}, S_j]), S_j \subseteq S, j = 1 \ldots k \right\}$.

Кожен вектор параметрів таким чином буде описувати один стан прогнозованого процесу на деякій підмножині вихідної вибірки. Дана постановка задачі робить припущення, що наявні дані описують кілька різних станів процесу, і кожен такий стан може бути описаний певним набором ваг. Однак кількість таких станів невідома - це завдання відповідного методу прогнозування.

б) функцію визначення кінцевого вектора параметрів θ_s для нових вхідних даних $\vec{x}_s$ виходячи з знайденої множини Θ векторів параметрів і самих вихідних даних: $\vec{\theta}_s = F(\vec{x}_s, \Theta, \mathrm{S})$;

такі, що мінімізують деяку оцінку помилки прогнозу $\hat{e} = e(F, \Theta, S)$.

III. Аналіз існуючих методів

Усі методи прогнозування можна розділити на 3 групи.

1. Глобальні методи. До цієї групи входить найбільша кількість методів: усі методи даної групи знаходять єдиний набір параметрів деякої прогнозуючої моделі, використовуючи всі наявні дані. З точки зору формалізованої постановки задачі методи в цій групі мають такі параметри:

а) $k = 1$ - використовується лише один вектор параметрів $\vec{\theta}_1$, при цьому $S_1 = S$ - значення цих параметрів знаходяться шляхом оптимізації деякого критерію L_1 на всіх вихідних даних;

б) в якості кінцевого вектора параметрів θ_s для нових вхідних даних $\vec{x}_s$ просто використовується вже знайдений вектор $\vec{\theta}_1$: $\vec{\theta}_s = \vec{\theta}_1$. При такому записі стає очевидно, що значення нових даних ніяк не використовуються при визначенні параметрів моделі.

Найбільш відомі методи в зазначеній групі: лінійна регресія, ARIMA, штучні нейронні мережі (ШНМ), МГУА.

2. Локальні методи, є протилежністю глобальним: замість однієї глобальної будується кілька локальних моделей, параметри для яких підбираються на деяких підмножинах вихідних даних. Дані методи враховують проблему непостійної поведінки прогнозованого процесу, але й мають свої недоліки:

- вони більш вимогливі до кількості навчальних даних - чим більше локальних моделей, тим більше потрібно даних;
- втрачається глобальна інформація - кожна локальна модель навчається тільки на своїх даних, що може привести до зайвої «спеціалізації» і перенавчання моделей.

Найбільш відомі методи в цій групі: локальна регресія, кластеризація з регресією, локальна апроксимація та ін..

3. Глобально-локальні методи. Методи з цієї групи намагаються об'єднати найкраще з обох підходів: всі

вихідні вхідні дані використовуються для отримання глобальної інформації про прогнозований об'єкт (часто званий трендом), в той час як деякі підмножини даних використовуються для знаходження локальних особливостей. У роботі [1] в якості прикладу глобально-локального методу названий метод MARS. Авторами також був розроблений підхід «перетворення» глобальних параметричних методів в глобально-локальні методи [2].

Суть підходу полягає в навчанні спочатку однієї глобальної моделі, після чого навчається множина локальних моделей з додатковим критерієм близькості параметрів локальних моделей до параметрів глобальної моделі; на стадії прогнозу просто вибираються параметри декількох «найближчих» локальних моделей, які потім усереднюються і застосовуються до нових входів (аналогічно деяким локальним методам). Завдяки такому підходу, а саме використанню критерію близькості параметрів локальної та глобальної моделей, при побудові локальних моделей не втрачається глобальна інформація та самим локальним моделям не потрібна велика кількість даних. Даний підхід був застосований до лінійної регресії, а отриманий метод названий «лінійна регресія з динамічними вагами» - ще один приклад глобально-локального методу.

Найбільш перспективними є глобально-локальні методи, так як вони позбавлені недоліків інших 2 груп:

- на відміну від глобальних методів, параметри прогнозуючої моделі залежать від значень вхідних змінних - тобто використовується локальна інформація про прогнозований процес;
- на відміну від локальних методів, не губиться глобальна інформація, що дозволяє уникнути перенавчання локальних моделей і знизити необхідну кількість навчальних даних.

Саме тому розроблений авторами підхід, що дозволяє переходити від глобальних до глобально-локальних методів, дає можливість отримувати більш якісну прогнозуючу модель (з точки зору мінімуму оцінки помилки прогнозу моделі), застосовуючи його до існуючих глобальних методів. Для демонстрації розглянемо глобально-локальний метод, отриманий шляхом застосування описаного підходу до глобального методу прогнозування з використанням нейронних мереж прямого поширення - найбільш відомої і застосовуваної архітектури нейронних мереж для вирішення задач прогнозування [3-5]. Основні переваги використання ШНМ для вирішення завдання прогнозування наступні:

- можливість навчання - нейронну мережу дуже просто адаптувати до нових даних - досить просто навчити її на них, при цьому результати попередніх навчань не втрачаються;
- можливість моделювати складні нелінійні залежності, функціональний вид яких заздалегідь невідомий.

IV. Аналіз алгоритму

Наведемо етапи отриманого глобально-локального методу прогнозування:

1. Навчається нейронна мережа прямого поширення з використанням множини усіх наявних прикладів $S = \{(\vec{x}_i, y_i) : \mathrm{i} = 1 \ldots \mathrm{n}\}$. У якості опції помилки мережі використовується стандартна середньоквадратична помилка:

$$E = \sum_{i=1}^{n} \left[net(\vec{x}_i) - y_i \right]^2$$

Ваги $\vec{\theta}^g$ даної мережі є інформацією про глобальний характер прогнозованого процесу. Питання вибору оптимальної структури нейронної мережі і оптимального алгоритму навчання в даній статті не розглядаються - автори використовували звичайний багатошаровий персептрон з одним прихованим шаром з 10 нейронів, з гіперболічним тангенсом у якості активаційної функції прихованого шару і лінійної функції в якості активаційної функції вихідного шару; для навчання використовувався алгоритм Rprop.

2. Виконується кластеризація множини прикладів S. В ідеалі, кожен кластер C_j повинен містити приклади, які найкращим чином описуються деяким єдиним набором ваг мережі $\vec{\theta}_j, j = 1 \ldots k$, де k - кількість кластерів, тобто, виконуючи кластеризацію, ми сподіваємося, що схожі приклади повинні описуватися схожими вагами моделі. Природно, якість прогнозування в такому випадку буде сильно залежати від використаного методу кластеризації. Для ілюстрації підходу в даній статті використовувався один з найбільш відомих методів кластеризації - k-means, з фіксованою кількістю кластерів - 10.

3. Навчаються k нейронних мереж - по одній на кожен кластер. Навчання виконується наступним чином:

а. Ваги кожної мережі ініціалізуються попередньо знайденими вагами $\vec{\theta}^g$.

б. Мережа net_j навчається тільки на прикладах з кластера C_j з використанням наступної функції помилки:

$$E_j = \gamma * \sum_{\vec{x}_i \in C_j} (net_j(\vec{x}) - y_i)^2 + \sum_{0_l \in \vec{0}_j} (\theta_l - \theta_l^g)^2 .$$

Таким чином, при навчанні ваги мережі будуть прагнути як зменшити помилку мережі на прикладах зі свого кластера, так і не віддалятися занадто сильно від початкових ваг $\vec{\theta}^g$ - тобто не буде втрачена інформація про глобальний характер процесу.

4. Після навчання локальних мереж, прогнозування для нових прикладів виконується наступним чином: для вхідного вектора $\vec{x}_s$ знаходиться відповідний йому

кластер C_h (шляхом знаходження найближчого центру кластера), і для прогнозу використовується мережа, навчена на прикладах цього кластера.

Описаний алгоритм було перевірено на наборі вибірок, що знаходяться у відкритому доступі в мережі Інтернет. У порівнянні з моделями, отриманими з використанням штучних нейронних мереж та лінійної регресії, середня похибка запропонованого методу була найменша, при чому середня похибка штучних нейронних мереж була більша на 24.4% а середня похибка лінійної регресії більша на 45.33%.

Висновки

Запропонований метод було порівняно з 2 іншими методами побудови прогнозуючих моделей на 11 різних наборах даних, і він мав найменшу середню середньоквадратичну помилку. Однак деякі питання у використанні описаного підходу залишаються невирішеними:

- оптимальний метод поділу наявних даних на кластери, що описують різні стани прогнозованого процесу;
- оптимальний спосіб об'єднання усіх знайдених параметрів моделі на етапі прогнозування.

Література References

[1] P. Lewis, and J. Steven, «Nonlinear modeling of time series using multivariate adaptive regression splines (MARS)», Journal of the American Statistical Association, 1991, 87, pp. 864–877.

[2] V. Sineglazov, O. Chumachenko, V. Gorbatiuk, «A method for building a forecasting model with dynamic weights», Восточно-Европейский журнал передовых технологий. - 2014. - № 2(4). - С. 4-8..

[3] F. A. Amir, I. S Samir, «A comparison between neural-network forecasting techniques – case study: river flow forecasting», IEEE Transactions on Neural Networks, 1999, 10(2), pp. 402–409.

[4] T. O. Blinova, «Analysis of possibility of using neural network to forecast passenger traffic flows in Russia», Aviation, 2007, 11(1), pp. 28–34.

[5] P. Kozik, J. Sęp, «Aircraft engine overhaul demand forecasting using ANN», Management and Production Engineering Review, 3(2), 2012, pp. 21–26.