УДК 004.432 004+ 004.942+519.254

С. Ю. Куницька, *к.т.н.,*
доцент кафедри інформаційної безпеки та комп'ютерної інженерії,
e-mail: kunitskaya33@gmail.com
Черкаський державний технологічний університет,
бульв. Шевченка, 460, м. Черкаси, 18006, Україна,

ТЕХНОЛОГІЯ ОБРОБКИ ІНФОРМАЦІЇ НОРМАЛІЗОВАНИХ СИСТЕМ

Для спрощення обчислювального процесу моделей будь-якої складності виникла необхідність розробити технологію обробки інформації у вигляді програмного алгоритму, який дозволяє обробляти, аналізувати та прогнозувати майбутній процес. Для перетворення вхідних даних у вихідну інформацію необхідно мати локальний алгоритм перетворення інформації, що грунтується на індуктивному методі синтезу цих алгоритмів. На першому етапі будь-якого ймовірнісного алгоритму синтезу описаний в статті алгоритм перетворення інформації - це представлення вхідних даних у вигляді перевизначеної системи умовних рівнянь. Надалі приведено системи до вигляду, який дозволяє визначити коефіцієнти системи. А потім введено поняття «нормалізації» системи, використовуючи метод найменших квадратів. Дослідження системи відбувається з точки зору її аргументів, що й дає змогу отримати необхідну навчену модель. Всі отримані модифікації навчених моделей перевірено на наявність похибки, що представляє собою середнєквадратичну похибку прогнозу для перевірки прогнозованої послідовності.

Ключові слова: *умовні рівняння, експериментальні точки, система, нормалізація, модель, похибка, прогнозування.*

.

Постановка проблеми. Побудова математичної моделі складного об'єкта або процеса є необхідністю для вирішення таких головних задач, як прогнозування або розпізнання образів. Під поняттям «математична модель» ми розуміємо систему рівнянь регресії, які використовуються для одноразового прогнозу майбутнього ходу процесу в складній системі.

Тобто, математична модель в задачах прогнозування може бути представлена як система рівнянь, що описує обєкт з певним числом експериментальних точок, необхідних для подальшого спостереження.

Аналіз останніх досліджень. Історичне виникнення мови програмування C# починає свій розвиток з 60-х років, де початком була мова B, що є представником імперативних мов програмування. Головною метою розробки цієї мови – це реалізація операційної системи UNIX. Наступним кроком розвитку була мова програмування C, що розроблена в 1972 році, де мова B розширилася за рахунок наочного використання типів, структур та нових операцій. Надалі, в 1984 році, введено поняття класу, як об'єкту даних та набуває розвитку мова C++ вже як об'єктно-орієнтована мова програмування. На сьогодні, мова C# ґрунтується на чітко побудованій компонентній архітектурі та реалізує механізми, що забезпечують принцип наслідування, побудову інтерфейсів, обробку виключних ситуацій і нитей (threads), а також безпеку динамічного завантаження кода при виконанні програми.

В цілому, корпорація Microsoft досить чітко реалізувала найважливіші задачі для програмування, що й відображені в мові програмування C#:

- компонентно-орієнтований підхід (що є характерним для ідеології Microsoft .NET);
- властивості, як засоби інкапсуляції даних;
- ситуаційна обробка (оператор try), делегати (delegate), індексатори – оператори індексу для звернення до елементів класу-контейнера (indexer), перезавантаженні оператори, оператори для обробки всіх елементів класів-колекцій (foreach – аналог Visual Basic), механізми для пере визначення типів (boxing і unboxing);
- атрибути, як засоби оперування метаданими в COM-моделі;
- прямокутні масиви;
- уніфікована система типізації, що відповідає ідеології Microsoft .NET в цілому.

Метою дослідження є нормалізація системи та отримання навченої моделі на основі розробленого програмного алгоритму, що зв'язує певний набір експериментальних точок за обраним діапазоном з введених раніше табличних даних, що зосереджені в Excel. Тобто описано програмний обчислювальний процес нормалізованих систем рівнянь, що будуються в залежності від кількості невідомих аргументів в опорному вигляді моделі полінома Колмогорова-Габора 2 ступеня з можливістю прогнозування поведінки майбутнього процесу.

Виклад основного матеріалу. Експериментальні точки, незалежно від рівня знань людини, містять в собі інформацію о різноманітних факторах [1]. Математична модель може бути побудована на різних наборах аргументів. Це ніяким чином не впливає на обєктивність прогнозуючої ситуації, але при цьому повинна бути забезпеченна достатня свобода вибору наступних рішень.

1. Перша і основна задача, що направлена на розробку та освітлена в цій статті - це написання програмного продукту на обраній мові програмування, що буде містити реалізацію обчислювального процесу за допомогою обрахунку полінома Колмогорова-Габора методом найменьших квадратів та приведення рівнянь до нормалізації з метою подальшого прогнозування [1].

Програмний продукт написано об'єктною мовою програмування C# з підтримкою інтегрованого середовища розробки Visual Studio Professional 2017 на платформі .NET, що має зручний інтерфейс користувача та містить необхідні для роботи кнопки управління.

Табличні значення за деякими двома рядками можуть бути як константи, так і введенні необхідні статистичні данні за необхідністю. Діапазон обчислення обирається також за необхідністю, при цьому обчислюються всі аргументи, що заносяться з таблиці для подальшого збереження інформації, аналізу та майбутнього прогнозування. Внесення даних та визначення діапазону експериментальних точок знаходяться в таблицях Excel, де зберігаються в різноматнітних файлах.

Програмний продукт максимально автоматизовано, що значно мінімізує втручання користувача, що й робить рішення поставлених задач найбільш універсальним. Дії користувача зводяться до введення з клавіатури певних експериментальних точок в таблицю, або обрання діапазону констант точок для переходу на наступний етап розробки. Чим більше буде введено експериментальних точок, тим складніше отримаємо майбутню навчену модель, але більш точніше отримаємо прогноз процесу [2]. Надалі необхідно обрати кількість невідомих аргументів по опорному вигляду полінома і на цьому втручання діяльності користувача обмежується. Далі обчислювальний процесс автоматизовано.

2. Опишемо та обгрунтуємо другу поставлену задачу, яка полягає в рішенні обчислювального процессу.

Різноманітні модифікації многорядного алгоритму відрізняються один від одного по виду опорної функції F. В алгоритмі з лінійними поліномами використовують часний опис виду:

$$y_k = a_0 + a_1 x_i + a_2 x_j, \qquad (1)$$

де $0 < i < m$, $0 < j < m$

Ускладнення моделі відбувається тільки за рахунок збільшення числа використаних аргументів, тобто в першому рядку селекції синтезуються моделі, що містять по два аргумента, на друдгому рядку синтезуються моделі з трьома або чотирма аргументами, на трерьому - може доходити до восьми аргументів.

Многорядні алгоритми при використанні нелінійних опорних функцій мають наступний вигляд:

$$y_k = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j \qquad (2)$$

Вони дозволяють отримати моделі практично будь-якої складності, тому що в кожному рядку селекції ступень полінома подвіюється. Число коефіцієнтів моделі при цьому може наближатися до нескінченності.

Особливість прогнозування – кожному наступному набору точок прогнозування передує ряд рішень попереднього набору точок обраних для написання системи умовних рівнянь, їх нормалізації та отримання навченной моделі в результаті.

В роботі розглянуто алгоритм отримання регрессійного, тобто за формальними ознаками відповідність першим двом загально відомим принципам самоорганізації [3]:

- дозволена повна свобода вибору як при включенні нових факторів в модель, так і при виключенні їх;
- граничне значення F-критерія є зовнішнім критерієм селекції, що обирається з таблиць із статистичними даними.

В свою чергу поняття самоорганізації щільно повязано з появою впорядкування в будь-якій системі. При цьому система повинна мати структуру, при дослідженні якої і будуть виникати критерії для її аналізу, прогнозування та поліпшенню процесу.

Надалі виникає необхідність в дослідженні системи з точки зору її поведінки, що й дає змогу в результаті отримати необхідну нам навченну модель. Це досягається за рахунок наступних етапів:

1. Спочатку обирається опорний вигляд полінома Колмогорова-Габора 2 ступеня [4]. Модель будь-якого динамічного процесу максимальної складності можна отримати із загального повного полінома (3):

$$\varphi = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k + ... \quad (3)$$

2. Далі вносимо данні в таблицю експериментальних точок, що й необхідні для подальшого обчислювального процессу;

3. Формуємо систему умовних рівнянь за обраним поліномом Колмогорова-Габора. Наприклад, для шести членів загальний вигляд полінома приймає наступний вид:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2 \quad (4)$$

4. Наступний етап – це отримання системи нормальних рівнянь за принципом: беремо кожне умовне рівняння та перемножуємо на коефіцієнти при першому невідомому, далі всі умовні рівняння додаємо. Для отримання другого нормалізованного рівняння берем умовне рівняння та перемножаємо на коефіцієнти при другому члені рівняння, знову додаємо. Цей процес повторюється до тих пір, поки кількість нормальних рівнянь не зрівняється з кількістю невідомих коефіцієнтів в системі умовних рівнянь.

5. Надалі нормалізована система рівнянь приводиться до матричного вигляду для подальшого її обрахування методом Гауса або методом найменьшого квадрата. Рішення нормалізованої системи рівнянь надає нам результат у вигляді отримання значень необхідних аргументів, які необхідні в подальшому для аналіза поведінки процесу та прогнозування.

Таким чином, ми отримали навчену модель, яку необхідно обовязково перевірити на наявність похибок [5].

Найбільш частіше використовують один з наступних критеріїв, що відповідає за визначення оптимальної моделі – це <u>критерій регулярності</u> [6, 7]. Він представляє собою середнє квадратичну похибку прогнозу для перевірки спрогнозованої послідовності за наступною формулою:

$$\delta^2 = \sum_{i=1}^{n_2} (y_{\phi i} - y_{\rho i})^2 / \sum_{i=1}^{n_2} y_{\phi i}^2 , \quad (5)$$

де $\boldsymbol{y_{\rho i}}$ – прогноз $\boldsymbol{i}$-го значення для перевірки прогнозованої послідовності; $\boldsymbol{y_{\phi i}}$ – фактичне значення цієї ж послідовності; $\boldsymbol{n_2}$ – об'єм послідовності

Тобто, отримана модель дозволяє получити найбільш точні кількісні значення вихідних змінних при однократному прогнозуванню майбутнього процесу. Але при збільшенні складності рівнянь регресії, такі як ступень полінома Колмогорова-Габора та число невідомих аргументів, похибка, що виникає в обчислювальному процесі поступово зменшує свої значення на всіх експериментальних точках [8, 9].

У випадку рівності кількості експериментальних точок числу коефіцієнтів рівняння в поліномі похибка майже наблизиться до нуля або буде дорівнювати нулю.

Висновки. Згідно сформованих задач розглянута та досліджена спрогнозована поведінка декількох математичних моделей, що отримані та навчені завдяки наступним крокам:

- в статті описан процес отримання моделі через поліном Колмогорова-Габора будь-якої складності із введенням діапазону деяких даних;

- отримано навчену модель, що задається зовнішніми критеріями, і яка вже досліджена на наявність похибок прогнозуючого процесу;

- для спрощення обробки моделей будь-якої складності розроблено технологію обробки інформації завдяки програмному алгоритму із зручним інтерфейсом користувача, що дозволяє не тільки автоматизувати процес обчислювання моделей, але й зберігати всю необхідну інформацію по отриманню навчених моделей різної складності таку як: кількість експериментальних точок, складність обраного поліному в залежності від невідомих аргументів, результати обчислення невідомих аргументів, модифікації навчених моделей та обчислення похибки прогнозуючого процесу [10].

Список літератури

1. Дайитбегов Д. М., Калмыкова О. В., Черепанов А. И. Программное обеспечение статистической обработки данных. М.: Финансы и статистика, 1984. 192 с.
2. Ивахненко А. Г. Долгосрочное прогнозирование и управление сложными системами. 1975. 312с.
3. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем. Киев: Наук. Думка,1982. 296 с.
4. Куницька С. Ю. Приведення системи умовних рівнянь до нормалізації. *Проблеми інформатизації:* тези доп. п'ятої міжнародн. наук.-техн. конф., Черкаси, 13-15 листопада 2017 року. Черкаси: ЧДТУ; Баку: ВА ЗС АР; Бельсько-Бяла: УТіГН; Полтава: ПНТУ, 2017. С. 60.
5. Флейшман Б. С., Брусиловский П. М., Розенберг Г. С. О методах математического моделирования сложных систем. *Системные исследования. Ежегодник*. М.: Наука, 1982. С. 65–79.
6. Ивахненко А. Г. Самообучающиеся системы распознавания и автоматического. Киев: Техника, 1969. 392 с.
7. Ивахненко А. Г., Зайченко Ю. П., Димитрова В. Д. Принятие решений на основе самоорганизации. М.: Сов. Радио, 1976. 275 с.
8. Редкозубов С. А. Статистические методы прогнозирования в АСУ. М.: Энергоатомиздат, 1981. 150 с.
9. Шеннон К. Е. Имитационное моделирование систем – искусство и наука. М.: Мир, 1978. 418 с.
10. Лисичкин В. А. Теория и практика прогностики. М.: Наука, 1972. 224 с.

References

1. Daiitbegov, D., Kalmykova, O., Cherepanov, O. (1984) Software for Statistical Data Processing. M.: Finance and Statistics. 192 p.
2. Ivakhnenko, A. (1975) Long-term forecasting and management of complex systems. Kiev: Technics. 312 p.
3. Ivakhnenko, A. (1982) Inductive method of self-organization of models of complex systems. Kiev: Science. Opinion. 296 p.
4. Kunytska, S. (2017) Bringing the system of conditional equations to normalization. *Problems of informatization:* theses of additional. fifth international. Sci.-Tech. Conf., Cherkasy, November 13-15, 2017. Cherkasy; Baku; Bielsko-Biala; Poltava. P. 60.
5. Fleischman, B., Brusilovsky, P., Rosenberg, G. (1982) About methods of mathematical modeling of complex systems. *System research. Yearbook.* M.: Science. P. 65–79.
6. Ivakhnenko, A. (1969) Self-learning recognition and automatic control systems. Kiev: Technics. 392 p.
7. Ivakhnenko, A., Zaychenko, Y., Dimitrova, V. (1976) Decision-making based on self-organization. M.: Sov. Radio. 275 p.
8. Rekzubov, S. (1981) Statistical methods of forecasting in ACS. M.: Energoatomizdat. 150 p.
9. Shannon, K. (1978) Imitation Modeling Systems – Art and Science. M .: Mir. 418 p.
10. Lisichkin, V. (1972) Theory and Practice of Prognostics. M.: Science. 224 p.

S. Kunytska, *Ph.D,*
associate professor of the department of information security and computer engineering
e-mail: kunitskaya33@gmail.com
Cherkasy State Technological University,
Shevchenko blvd., 460, Cherkasy, 18006, Ukraine

TECHNOLOGY OF INFORMATION PROCESSING OF NORMALIZED SYSTEMS

To simplify the computational process of models of any complexity, there was a need to develop a technology for processing information in the form of a software algorithm that allows you to process, analyze and predict the future process. To convert input data into source information, it is necessary to have a local information conversion algorithm based on the inductive method of synthesizing these algorithms. In the first stage of any probabilistic synthesis algorithm, the algorithm for information transformation is described in the article as the representation of the input data in the form of a predefined system of conditional equations. In the following, the system is presented in a form that allows you to determine the coefficients of the system. And then the concept of "normalization" of the system is introduced, using the method of least squares. The system's research takes place from the point of view of its arguments, which makes it possible to obtain the necessary trained model. All received modifications of the trained models were checked for an error, which is the mean square prediction error for checking the predicted sequence.

Keywords: *conditional equations, experimental points, system, normalization, model, error, forecasting*

Рецензенти: С.В. Голуб, д.т.н., професор,
Т.О. Прокопенко, д.т.н., доцент