УДК 519.711

Артемчук В.О.

МОДЕЛЬ ВИЗНАЧЕННЯ АНАЛІТИЧНОЇ ЗАЛЕЖНОСТІ ЗА ДИСКРЕТНИМ ЗАДАННЯМ ФУНКЦІЇ З ВИКОРИСТАННЯМ ТЕОРІЇ ЛІНІЙНОГО ПРОГРАМУВАННЯ

Вступ

Більшість параметрів (факторів), що визначають реалії нашого життя можуть бути виміряні (обчислені) та подані у вигляді деякої числової послідовності за m періодів (спостережень, експериментів):

$$y_1, y_2, ..., y_m \tag{1}$$

залежать від n деяких інших (називатимемо їх незалежними) факторів, для яких також є відомими m наборів числових значень:

$$x_{1i}, x_{2i}, ..., x_{ni}, \ i = \overline{1, m} \tag{2}$$

Часто для моделювання відповідного взаємозв'язку між залежним фактором y та визначеними незалежними факторами x_j, $j = \overline{1, n}$ та/або прогнозування значення y_p в залежності від значення незалежних факторів:

$$x_{1p}, x_{2p}, ..., x_{np} \tag{3}$$

необхідне вирішення задачі визначення аналітичної залежності деякої (залежної) функції:

$$y = f(x_1, x_2, ..., x_n) \tag{4}$$

від n незалежних змінних (факторів) $x_1, x_2, ..., x_n$ за відомими її значеннями в m точках:

$$y_i = y(x_{1i}, x_{2i}, ..., x_{ni}), \ i = \overline{1, m} \tag{5}$$

Тому вирішення задачі визначення аналітичної залежності (4)-(5) є актуальним.

У даній роботі розглянуті основні моменти створення моделі для вирішення задачі визначення аналітичної залежності (4)-(5) та її програмної реалізації. Обґрунтовано вибір методів та інструментів для вирішення поставленої задачі. Наведено приклади роботи створених на основі розробленої моделі програмних модулів.

Методи дослідження

Відповідно до поставленої задачі було проведено дослідження теоретичних підходів до розв'язку задачі визначення аналітичної залежності (4)-(5) та інструментів для їх програмної реалізації. Проведено перевірку отриманих результатів на практиці.

Вирішення задачі

Побудуємо оптимізаційну модель для вирішення задачі визначення аналітичної залежності (4)-(5). Функція цілі, що мінімізує сумарне відхилення фактичних значень функції y від їх значень, що отримані із шуканої аналітичної залежності (4):

$$F = \sum_{i=1}^{m} (X_i^{'} + X_i^{''}) \to \min \tag{6}$$

за наступних обмежень:

$$\begin{cases} f(x_{11}, x_{21}, ..., x_{n1}) + X_1^{'} - X_1^{''} = y_1 \\ f(x_{12}, x_{22}, ..., x_{n2}) + X_2^{'} - X_2^{''} = y_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ f(x_{1i}, x_{2i}, ..., x_{ni}) + X_i^{'} - X_i^{''} = y_i \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ f(x_{1m}, x_{2m}, ..., x_{nm}) + X_m^{'} - X_m^{''} = y_m \end{cases} \tag{7}$$

$$\begin{cases} X_1^{'} \geq 0 \\ X_2^{'} \geq 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X_i^{'} \geq 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X_m^{'} \geq 0 \end{cases} \tag{8}$$

$$\begin{cases} X_1^{''} \geq 0 \\ X_2^{''} \geq 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X_i^{''} \geq 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X_m^{''} \geq 0 \end{cases} \tag{9}$$

де:

$X_i^{'}$ – модуль від’ємного відхилення фактичного рівня ряду від тренду;

$X_i^{''}$ – додатне відхилення фактичного рівня ряду від тренду;

$y = f(x_1, x_2, ..., x_n)$ – шукана аналітична залежність;

y_i – фактичні значення функції y;
n – кількість незалежних змінних в (4);
m – кількість статистичних даних (5) для визначення аналітичної залежності (4).

Перепишемо (7)-(9) у вигляді:

$$\begin{cases} f(x_{1i}, x_{2i}, ..., x_{ni}) + X_i^{'} - X_i^{''} = y_i \\ X_i^{'} \geq 0 \\ X_i^{''} \geq 0 \end{cases} \tag{10}$$

$$i = \overline{1,m}$$

Функцію цілі (6) можна подати і в наступних модифікаціях:

$$F = \sum_{i=1}^{m} p_i (X_i^{'} + X_i^{''}) \to \min \tag{11}$$

або

$$F = \sum_{i=1}^{m} c_i (X_i^{'} + X_i^{''}) \to \min \tag{12}$$

де:
p_i – достовірність наявних даних за *i*-тий період (чи отриманих даних від і-того спостереження чи експерименту). $p_i \in (0;1]$. У випадку $p_i = 0$ і-ті обмеження можуть бути виключені з (10);
c_i – числовий коефіцієнт вагомості наявних даних за і-тий період (чи отриманих даних від і-того спостереження чи експерименту) $c_i > 0$. У випадку $c_i = 0$ і-ті обмеження можуть бути виключені з (10).

Зауважимо, що (6) є частинним випадком (11) при:

$$p_i = 1, i = \overline{1,m} \tag{13}$$

або (12) при:

$$c_i = 1, i = \overline{1,m} \tag{14}$$

Внаслідок розв'язання оптимізаційної моделі (6), (10) або (10)-(11) або (10), (12) буде отримана шукана аналітична залежність (4).

Наприклад шукатимемо залежність (4) від двох змінних ($n = 2$) у вигляді:

$$y = f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_1^2 + a_4 \sin(x_1) + a_5 x_2 + a_6 x_2^2 + a_7 x_2^3 + a_8 e^{x_2} \tag{15}$$

використовуючи модель (6), (10) за наявними даними для 35 спостережень ($m = 35$).

Тоді (6), (10) матиме вигляд:

$$F = \sum_{i=1}^{35} (X_i^{'} + X_i^{"}) \rightarrow \min \tag{16}$$

$$\begin{cases} a_0 + a_1 x_{1i} + a_2 x_{1i}^2 + a_4 \sin(x_{1i}) + a_5 x_{2i} + a_6 x_{2i}^2 + \\ \qquad + a_7 x_{2i}^3 + a_8 e^{x_{2i}} + X_i^{'} - X_i^{"} = y_i \\ X_i^{'} \geq 0 \\ X_i^{"} \geq 0 \end{cases} \tag{17}$$

$$i = \overline{1,35}$$

Отримана таким чином задача містить $k = 9$ коефіцієнтів у шуканій залежності, тобто, знайшовши їх числове значення ми отримуємо шукану залежність.

Отримана задача (16)-(17) є задачею лінійного програмування, оскільки містить лінійну функцію цілі (16) та лінійні відносно невідомих змінних

$$a_j, j = \overline{0, k-1} \tag{18}$$

та

$$X_i^{"},\ X_i^{"},\ i = \overline{1, m} \tag{19}$$

обмеження - рівняння та нерівності з (17).

Отримана задача (16)-(17) легко приводиться до канонічного вигляду, що є зручним для застосування симплекс-методу. Оскільки на змінні з (19) накладено умови невід'ємності, а на змінні з (18) – ні, то робимо заміну:

$$a_j = a_j^{'} - a_j^{"}, a_j^{'} \geq 0,\ a_j^{"} \geq 0,\ \ j = \overline{0, k-1} \tag{20}$$

За базисні змінні для першого обмеження з (17) приймаємо $X_i^{"}$.

Якщо:

$$y_i \geq 0,\ i = \overline{1, m} \tag{21}$$

то приведення задачі (16)-(17) закінчено.

Якщо умова (21) не виконується, то знаходиться:

$$y_{\min} = \min\{y_i\},\ i = \overline{1, m} \tag{22}$$

Після чого робиться заміна:

$$y_i = y_i - y_{\min},\ i = \overline{1, m} \tag{23}$$

Та після розв'язання задачі враховуємо (23) шляхом додавання до вільного члену a_0 значення $y_{\min}$:

$$a_0 = a_0 + y_{\min} \tag{24}$$

Отримана таким чином із (16)-(17) задача завжди має розв'язок (оскільки множина її розв'язків завжди не порожня), та може бути розв'язана симплекс-методом.

На основі висновків, зроблених в [2], для часткової програмної реалізації вказаної моделі обрано середовище програмування Borland C++ Builder 6.0.

Перевіримо адекватність результатів, отриманих за допомогою розробленої моделі, використовуючи розроблений програмний продукт.

Розглянемо однофакторну залежність:

$$y = f(x) = 4 + 2x + 0{,}05x^5 - 3\ln x - 1{,}5\cos x \qquad (25)$$

Побудуємо наступну таблицю її числових значень:

Таблиця 1. Значення функції (25)

X	Y	Y (заокруглене)
0,82	5,230557997	5,2306
2,02	8,263696065	8,2637
2,22	9,650399867	9,6504
2,52	12,56792645	12,568
2,62	13,82373547	13,824
2,82	16,86982305	16,87
2,92	18,70269868	18,703
3,42	31,98710706	31,987
3,72	44,3740751	44,374
3,82	49,45829282	49,458
4,62	114,0269703	114,03
4,92	152,8945412	152,89
5,02	168,1456364	168,15
5,22	202,5394254	202,54
5,32	221,8420935	221,84
5,52	265,0818058	265,08
6,12	438,5918578	438,59
6,42	555,0878754	555,09
7,02	863,5065877	863,51
7,32	1062,717656	1062,7
7,62	1297,331695	1297,3
7,72	1384,168328	1384,2
8,22	1891,072987	1891,1
8,52	2260,289531	2260,3
8,82	2685,123833	2685,1
8,92	2840,130191	2840,1
8,92	2840,130191	2840,1
9,12	3171,638713	3171,6
9,72	4356,17023	4356,2
11,02	8144,825914	8144,8

11,32	9312,870808	9312,9

При використанні точних значень (колонки 1 і 2 таблиці 1) для визначення залежностей отримаємо наступний результат (див. рис. 1):

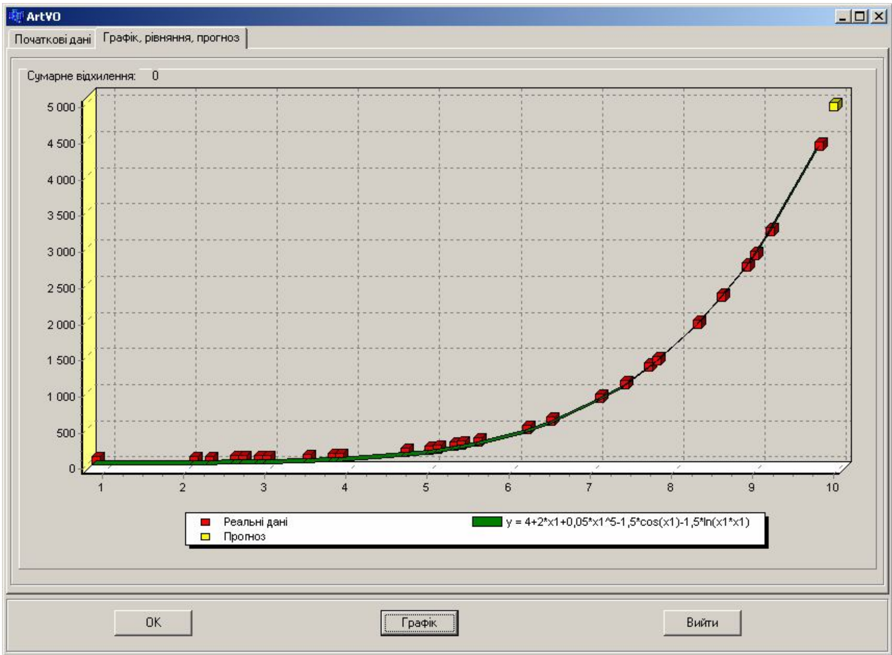

Рис. 1. Отриманий результат – аналітична функція залежності

$$y = 4 + 2x + 0{,}05x^5 - 1{,}5\ln x^2 - 1{,}5\cos x = 4 + 2x + 0{,}05x^5 - 3\ln x - 1{,}5\cos x$$

Як бачимо, у цьому випадку модель та її програмна реалізація дають точний результат.

Звісно, якщо залежність (4) точного вираження не має, або не враховані всі фактори, що впливають на залежну змінну, то результатом роботи програмної реалізації моделі буде „найкраща підігнана” функція.

Це легко перевірити, коли використати неточні значення функції (25)

При використанні неточних (наближених) значень (колонки 1 і 3 табл. 1) для визначення залежностей отримаємо результат, зображений на рис. 2.

Такими ж були результати інших проведених експериментів з функціями однієї, двох, трьох та чотирьох змінних.

Додамо, що створений програмний продукт може використовуватися для визначення прогнозних значень функції y за побудованою залежністю (функцією).

На даний момент реалізовано у вигляді програмного модулю процес побудови залежностей у вигляді суми лінійних комбінацій функцій синуса, косинуса, натурального логарифму, експоненти та поліному для кожної незалежної змінної із можливістю вибору користувачем функцій вказаних лінійних комбінацій для кожної незалежної змінної. Це може бути дуже корисним, коли наперед відомо деякі закономірності: наприклад відомо, що y поліноміально залежить від x_1, експоненціально від x_2, тощо.

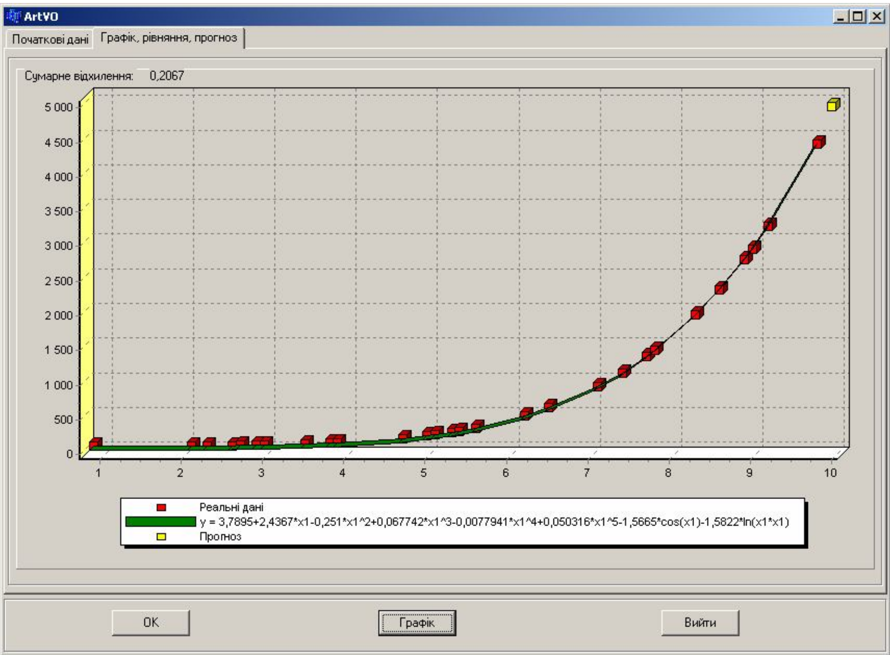

Рис. 2. Отриманий результат – аналітична функція залежності

$$y = 3,79 + 2,43x - 0,25x^2 + 0,07x^3 - 0,008x^4 + 0,05x^5 - 1,58\ln x^2 - 1,57\cos x \approx$$
$$\approx 4 + 2x + 0,05x^5 - 3\ln x - 1,5\cos x$$

Результати роботи

Результатом проведеної роботи є побудова оптимізаційної моделі для вирішення задачі визначення аналітичної залежності (4)-(5).

Доведено, що внаслідок розв'язання отриманої оптимізаційної моделі (6), (10) або (10)-(11) або (10), (12) буде отримана шукана аналітична залежність (4).

Показано приклад побудови моделі для відшукання конкретної залежності на прикладі (15).

Наведено алгоритм приведення моделі до виду, що є зручним для застосування симплекс-методу за допомогою (20), (22), (23).

На основі даної моделі було створено відповідний програмний продукт, за допомогою якого було здійснено перевірку ефективності моделі на практиці.

Надалі планується розвиток даної моделі та даного програмного продукту.

Висновки

В роботі обґрунтовано актуальність задачі визначення аналітичної залежності (4)-(5). Поставлено відповідну задачу, створено модель для її вирішення та здійснено її програмну реалізацію. Виходячи з особливостей поставленої задачі, для її програмної реалізації було обрано середовище програмування Borland C++ Builder 6.0. Отримані результати перевірені на практиці. Створену програмну реалізацію даної моделі планується використати в системі, побудова якої описана в [1], [2], [5].

1. *Артемчук В.О.* „База даних еколого-енергетичного моніторингу: проектування та створення” // Збірник наукових праць ІПМЕ ім. Г.Є. Пухова НАН України 2008 р.
2. *Артемчук В.О.* „Інтеграція бази даних еколого-енергетичного моніторингу в програмний додаток” // Збірник наукових праць ІПМЕ ім. Г.Є. Пухова НАН України 2009 р.
3. *Архангельський А. Я.* Программирование в C++ Builder 6 — М.: «Бином», 2002. – С. 1152.
4. *Джаррод Холингворт, Боб Сворт, Марк Кэшмэн, Поль Густавсон* Borland C++ Builder 6. Руководство разработчика / Borland C++ Builder 6 Developer's Guide. – М.: «Вильямс», 2004. — С. 976.
5. *Яцишин А.В., Артемчук В.О.* „Формування вибірки з бази даних еколого-енергетичного моніторингу” // Матеріали Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій» ISDMIT’2009р., 18-21 травня 2009 р., Євпаторія. 2009.