

УДК 004.94

С.В. Голуб, канд. техн. наук, доц.

ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ “РОЗСТАНОВКА АКЦЕНТІВ” ПРИ ІНФОРМАЦІЙНОМУ МОДЕЛЮВАННІ В МОНІТОРИНГОВИХ СИСТЕМАХ

Черкаський національний університет ім. Богдана Хмельницького

Запропоновано технологію підвищення стійкості індуктивного моделювання об'єктів моніторингу довкілля. Експериментально підтверджено, що застосування нової технології дозволяє підвищувати стійкість інформаційної моделі до 30 %.

The technology of increase of stability of inductive modeling of objects of an environment is offered. It is experimentally confirmed, that application of new technology allows to raise stability of information model up to 30 %.

Постановка проблеми

Удосконалення системи моніторингу стану довкілля Черкаської області передбачає побудову інформаційної моніторингової системи на базі існуючих підрозділів обласної санітарно-епідеміологічної станції (СЕС) та обласного управління екології та природних ресурсів згідно з договором № 2 від 10 березня 2004 р. на створення науково-технічної продукції “Створення реєстру забруднюючих довкілля викидів та їх переносу в області” на підставі розпорядження голови Черкаської облдержадміністрації від 25.05.2000 р. № 189 «Про координаційну раду зі створення реєстру забруднюючих довкілля викидів та їх переносу в області» та звернення Інституту медицини праці Академії медичних наук України від 17.04.2000 № 01/255.

Моніторинг організовується як комплекс заходів з моделювання об'єктів довкілля на підставі евристичної системи спостереження (ЕСС) [1].

Основна мета цих заходів – отримання моделей, які дозволяють прорахувати наслідки того чи іншого впливу на довкілля для поліпшення стану здоров'я населення та забезпечення комфортних умов його існування через збереження природного середовища [2].

Основою моделювання об'єктів довкілля за допомогою ЕСС є індуктивний метод, реалізований за алгоритмами методу групового врахування аргументів (МГВА) [3, с. 38]. Модель отримують у вигляді аналітичної залежності однієї з характеристик досліджуваного об'єкта від інших його характеристик, які використані як параметри моделювання.

Генерація інформаційної моделі організована як багатоетапний процес селекції моделей на основі евристичної самоорганізації [4].

Визнаним способом підвищення якості моделей є використання принципу зовнішнього критерію при формуванні первинного опису об'єкта спостереження.

Первинний опис містить числові значення характеристик об'єкта, отриманих експериментально в одних і тих самих умовах, що і значення модельованої функції. На основі цих даних формується матриця спостережень. Принцип зовнішнього критерію передбачає розбиття первинного опису як мінімум на дві частини: навчаючу та перевірну. Для оцінки якості моделей використовують експериментальні дані, які не застосовували для їх навчання [5].

Для розбиття таблиці експериментальних вхідних даних на навчаючу та перевірну послідовності використовують кілька способів [5]. Таблиця вхідних даних ділиться на дві частини або рівні, або обсяг навчаючої послідовності більша за обсяг перевірної. В деяких випадках рядки з непарним номером формують навчаючу послідовність, з парним – перевірну.

Перед розбиттям застосовують процедуру ранжування – розміщують рядки таблиці за величиною дисперсії модельованої функції від значення її середнього арифметичного [6].

Послідовності пропонують формувати за результатами моделювання [7], організовуючи перебір способів формування цих послідовностей за критерієм незміщеності, враховуючи результати аналізу функції розподілу зовнішнього критерію [8].

У праці [4] описано спосіб розділення вхідних даних на навчаючу та перевірну послідовність за ознакою мінімуму рядів селекції багаторядного алгоритму МГВА. При ранжуванні вхідних даних за дисперсією було експериментально виявлено існування мінімуму рядів селекції при зміні співвідношення кількості точок навчаючої та перевірної послідовності. Чим менше рядів селекції та простіша модель, тим вона достовірніша.

Ефективним вважається розбиття первинного опису об'єкта за критерієм мінімізації дисперсії похибки прогнозу за моделлю оптимальної складності [9].

В умовах багаторівневого моніторингу, коли необхідно застосовувати технології індуктивного моделювання для дослідження об'єктів мікро-, макро- та метарівнів [10] за однотипними алгоритмами невраховані в моделі фактори призводять до значних відхилень модельованих значень функції від її експериментальних значень. Крім того, технологія інформаційного моделювання об'єктів довкілля, яка припускає можливість моделювання на всіх рівнях моніторингу за однотипними алгоритмами, має свої особливості, що відрізняють її від моделювання об'єктів тільки на одному із рівнів.

У таких умовах традиційні технології формування зовнішнього критерію недостатньо ефективні, тому необхідно провести додаткові дослідження процесу формування зовнішнього критерію якості інформаційних моделей.

Мета цієї роботи – дослідження ефективності запропонованої автором технології формування перевірної послідовності первинного опису об'єкта при індуктивному моделюванні в моніторингових системах.

Гіпотеза дослідження

Для підвищення якості отриманої моделі необхідно посилювати зовнішній критерій перенесенням спостережень, значення функції яких модельовані з найбільшою похибкою, із навчаючої послідовності в перевірку. Для перевірки цієї гіпотези проведено модельний експеримент.

Умови проведення модельного експерименту

Для експерименту використаний багаторядний алгоритм МГВА [3].

Спостереження розбиті на дві рівні частини, перша з яких є навчаючою послідовністю, друга – перевіркою. Критерій інцухту визначався за середньою похибкою моделей ряду.

Об'єктом моніторингу є залежність захворюваності населення від концентрації шкідливих речовин у повітрі Черкаської області.

Результати спостережень отримано під час створення реєстру забруднюючих довкілля викидів та їх перенесення в області за даними Черкаської обласної СЕС та обласного управління екології та природних ресурсів.

Об'єктом моделювання була залежність рівня захворюваності цукровим діабетом населення районних центрів Черкаської області в 1999–2002 рр. від концентрації шкідливих речовин у повітрі житлової зони та інших параметрів:

- окислів азоту;
- сірчистого ангідриду;
- окису вуглецю;
- вуглеводнів;
- діоксиду марганцю;
- аміаку;
- сажі;
- толуолу;
- бутилацетату;
- ацетону;
- легких органічних сполук;
- оксиду кальцію;
- ксилолу;
- етил-ацетату;
- золи ТЕС;
- бензину;
- фенолу.

Для налаштування генератора інформаційної моделі експериментально визначався оптимальний вигляд опорної моделі.

Оптимальною визнавали модель, яка одночасно мала максимальну точність та стійкість моделювання. При цьому показником точності моделі є мінімум абсолютної похибки моделювання.

Стійкість моделі оцінювали за мінімумом суми квадратів похибок (табл. 1).

Таблиця 1

Результати дослідження оптимального вигляду опорної моделі

Номер моделі	Вигляд початкової моделі	Абсолютна похибка моделі захворювань 10 000 населення	Сума квадратів похибок захворювань 10 000 населення
1	$Y=a_0+a_1x_1+a_2x_2;$	21,05	92642,21
2	$Y=a_0+a_1x_1+a_2x_2+a_3x_1x_2$	18,44	82451,14
3	$Y=a_0+a_1x_1+a_2x_2+a_4x_1^2+a_5x_2^2$	18,26	83720,10
4	$Y= a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_1^2+a_5x_2^2$	17,24	73798,49
5	$Y= a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_1^2+a_5x_2^2+a_6x_1^2x_2+a_7x_1x_2^2+a_8x_1^2x_2^2$	17,20	72426,80

За результатами експерименту оптимальною є опорна модель 5.

Після визначення оптимальних налаштувань класифікатора та моделювання визначаємо спостереження навчаючої послідовності, які мають найбільшу похибку моделювання (табл. 2).

Далі зазначені моделі перенесли в перевірну послідовність (табл. 3).

Сума квадратів похибок кожної із п'яти отриманих моделей менша порівняно з цим показником моделі 5 (табл. 1).

Тобто застосування технології “Розстановка акцентів” призводить до отримання більш стійкої інформаційної моделі.

Достатньо перенести в перевірну послідовність спостереження 1 навчаючої послідовності, яке має найбільшу похибку моделювання (табл. 3, модель 4).

Висновки

Застосування технології “Розстановка акцентів” дозволяє підвищити точність моделювання. Стійкість інформаційної моделі при цьому збільшується на 30 %. Це підтверджує ефективність запропонованої технології.

Максимальний ефект при застосуванні технології досягається при перенесенні з навчаючої послідовності в перевірну одного спостереження з найбільшою похибкою моделювання.

Таблиця 2

Спостереження навчаючої послідовності з найбільшою похибкою моделювання

Номер моделі	Номер спостережень, які мають найбільшу похибку при застосуванні моделі	Абсолютна похибка моделі захворювань 10 000 населення	Відносна похибка моделі, %
1	12	55,49	-17,80
2	23	28,28	-11,11
3	24	42,33	-15,71
4	25	78,44	48,00
5	26	64,08	36,39
6	27	42,85	21,67
7	37	71,71	-22,11
8	45	56,32	28,67

Таблиця 3

План модельного експерименту та його результати

Номер моделі	Номер спостереження, перенесеного в перевірну послідовність	Абсолютна похибка моделі захворювань 10 000 населення	Сума квадратів похибок захворювань 10 000 населення
1	12, 23, 24, 25, 26, 27, 37, 45	21,427	57758,59
2	25, 26, 27,	19,536	51368,22
3	25	17,073	51090,80
4	25, 26	19,81	62550,25

Література

1. Голуб С.В. Моделирование мониторинговых процессов эвристической системой наблюдений // Электронное моделирование. – К. – Т. 26, № 5. – С. 55–65.
2. Самотуга В.В., Загородній М.В., Терещенко В.Г., Голуб С.В. Використання регіональних реєстрів шкідливих речовин (РВПЗ) для зниження захворюваності населення // Тез. доп. II з'їзду токсикологів України. – К., 2004. – С. 82–83.
3. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. – К.: Наук. думка, 1981. – 296 с.
4. Ивахненко А.Г. Системы эвристической самоорганизации в технической кибернетике. – К.: Техніка, 1971 – 372 с.
5. Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д. Принятие решений на основе самоорганизации. – М.: Сов. радио, 1976. – 280 с.

6. Павлов О.А. Критерий ранжирования для порогового самовідбору змінних в алгоритмах МГУА // Автоматика. – 1969. – № 4. – С. 89–91.
7. Висоцький В.М. Про найкращий поділ вхідних даних в алгоритмах МГУА// Автоматика. – 1976. – № 3. – С. 71–74.
8. Юрачковський Ю.П., Горшков А.Н. Оптимальное разбиение исходных данных на обучающую и проверочную последовательности на основе анализа функции распределения критерия. //Автоматика. – 1980. – № 2. – С. 5–9.
9. Степашко В.С., Кондрашова Н.В. Исследование способов генерации вариантов разбиения выборки в алгоритмах МГУА// Пр. I Міжнар. конф. з індуктивного моделювання. – Л. 20–25 трав. 2002. – Т. 1, ч.1. – С. 90–94.
10. Голуб С.В., Джолос В.М. Використання евристичної системи спостереження для прогнозування післяінфарктних ускладнень // Вісн. Житомир. держ. технол. ун-ту. Техн. науки. – Житомир: ЖДТУ :Т. 2, № 4 (31). – С. 109–114.