

УДК 004.9+504.064.3

С.В. Голуб, д.т.н.

ФОРМУВАННЯ ДОДАТКОВИХ ОЗНАК МАСИВІВ ВХІДНИХ ДАНИХ В ІНФОРМАЦІЙНИХ СИСТЕМАХ КРИЗОВОГО СОЦІОЕКОЛОГІЧНОГО МОНІТОРИНГУ

Черкаський національний університет імені Богдана Хмельницького

З метою підвищення інформативності масиву вхідних даних та покращення характеристик індуктивних моделей запропоновано новий метод формування додаткових показників. Від існуючих цей метод відрізняється тим, що на першому етапі «формування взаємодій» проводиться кластеризація показників, після чого взаємодії формуються між показниками одного кластера. Експериментально підтверджено ефективність застосування нового методу в індуктивних алгоритмах синтезу моделей в системах кризового соціоекологічного моніторингу.

Ключові слова: синтез моделей, масив вхідних даних, додаткові показники, кризовий моніторинг.

Вступ

Інформаційні системи кризового соціоекологічного моніторингу створюються з метою виявлення характеристик захворюваності населення в наслідок техногенних аварій, терористичних актів та інших подібних подій.

Кризовий моніторинг вимагає приймати рішення про компенсаційні заходи за короткі проміжки часу, тому вимоги до часу реакції системи на зміну властивостей вхідних даних значно жорсткіші порівняно з подібними вимогами до існуючих автоматизованих систем багаторівневого соціоекологічного моніторингу [1]. Перетворення інформації в системах цього типу реалізується шляхом послідовного використання моделей об'єктів моніторингу, поєднаних в ієрархічну структуру. Якість моделей періодично контролюється шляхом використання масивів дійсних даних, отриманих в результаті спостереження за об'єктом.

Для забезпечення відповідності цих інформаційних систем вимогам кризового моніторингу необхідно забезпечити відображення властивостей об'єктів довкілля після кризової події в структурі нових моделей, які формують структуру підсистеми перетворення інформації.

Постановка задачі

В умовах різкого збільшення концентрації шкідливих речовин в об'єктах довкілля та підвищення рівня техногенних випромінювань об'єкти моніторингу набувають нового стану, який лише частково відображений в масиві даних первинного опису. За цих умов адекватність моделей, які були попередньо синтезовані і відображають залежність показників захворюваності населення від характеристик об'єктів довкілля, знижується.

Обмеження часу реакції системи виключає можливості реєстрації значення додаткових показників стану об'єктів моніторингу, тому підвищення адекватності нових моделей може досягатись за рахунок використання інформації, яку вже містить масив даних первинного опису. Необхідно підвищити адекватність кризових моделей шляхом збільшення кількості існуючих показників масиву вхідних даних (МВД).

Синтез моделей об'єктів кризового моніторингу відбувається за МВД первинного опису:

$$X = \{x_1, x_2, \dots, x_n\} \quad (1)$$

де x_1, x_2, \dots, x_n – показники техногенних випромінювань та концентрації шкідливих речовин у воді, повітрі та продуктах харчування.

Моделі синтезуються індуктивним методом, наприклад за багаторядним алгоритмом МГУА [2], у вигляді функціональних залежностей:

$$Y = Y(X), \quad Y = \{y_1, y_2, \dots, y_m\} \quad (2)$$

де y_1, y_2, \dots, y_m – показники захворюваності населення.

Якість моделей оцінюється за критерієм регулярності [2] – розраховується середнє квадратичне відхилення результатів моделювання на екзаменаційній послідовності даних, які не використовувались в процесі синтезу моделі:

$$S_y = \sqrt{\frac{\sum_{i=1}^k (y_i - y_i^*)^2}{k}}, \quad (3)$$

де y_i – розраховане значення i -го показника захворюваності населення; y_i^* – дійсне значення i -го показника захворюваності населення; k – кількість спостережень екзанаційної послідовності.

Ставиться задача підвищення адекватності моделей об'єктів кризового моніторингу шляхом збільшення кількості показників МВД за рахунок формування нових ознак:

$$Y = Y(x_1, x_2, \dots, x_n, z_1(X), z_2(X), \dots, z_l), S_y \rightarrow \min \quad (4)$$

Нові ознаки МВД формуються у вигляді функції від вже існуючих показників стану об'єкта моніторингу:

$$z_i(x) = z_i(x_1, x_2, \dots, x_n), i = 1, \dots, l \quad (5)$$

де l – кількість додатково сформованих параметрів.

Аналіз останніх досліджень та публікацій.

Якість моделей, отриманих за технологією індуктивного моделювання залежить від кількості корисної інформації, яку містить МВД та від „різноманітності” алгоритму синтезу моделей. Необхідна адекватність моделювання досягається за умови, коли кількість корисної інформації первинного опису об'єкта моніторингу досягає межі інформаційної достатності (МІД) [3]. Розробка нового методу вимірювання корисної інформації в індуктивних моделях об'єктів СЕМ [4] дозволяє визначити МІД та оцінити потенційну можливість синтезу адекватної моделі існуючими засобами. Випадки недостатньої різноманітності методів синтезу однорівневих моделей описані як обмеження однорівневості. Для подолання цих обмежень запропоновано застосовувати методи багаторівневого моделювання [5].

Одним із способів підвищення кількості корисної інформації є зниження суміщеності сигналів первинного опису які визначають структуру моделі [3]. Суміщеністю оцінюється кількість однакової інформації, яку несуть сигнали. Виявлено, що при зниженні рівня суміщеності сигналів підвищується адекватність та врівноваженість моделей [6].

Підвищення інформативності МВД за рахунок збільшення кількості несумішених показників досягалась шляхом застосування результатів кореляційного аналізу [7].

В [8] описано спосіб «формування взаємодій», де кількість показників МВД збільшується шляхом перемноження значень показників, розміщених у відповідних стовпцях.

Формулювання цілі статті

Метою даної роботи є дослідження методу формування додаткових показників МВД в процесі синтезу моделей об'єктів кризового соціоекологічного моніторингу.

Виклад основного матеріалу

Під час первинної обробки результатів спостережень кількість інформації в МВД не змінюється. В процесі синтезу моделі інформація ПО відображається в структурі моделі. Від адекватності та повноти відображення інформації в моделі залежить скільки її стає доступною при формуванні рішень із управління техногенним навантаженням довкілля.

При незмінності кількості інформації в МВД та проблемності отримання додаткових даних єдиним способом підвищити якість моделей є розробка методів підвищення різноманітності методів синтезу моделей.

Керуючись принципами зниження суміщеності інформації в показниках, які чисельно відображають характеристики об'єкта, була сформульована гіпотеза про те, що для генерації несумішених показників необхідно використовувати інформацію про належність показників, що генерують сигнали, до спільних кластерів.

Для підтвердження сформульованої гіпотези був проведений модельний експеримент. Об'єктом моделювання була залежність рівня захворюваності населення Черкаської області на злоякісні утворення від концентрації шкідливих речовин в повітрі жилої зони. В таблиці 1 подані показники масиву вхідних даних.

Для кластеризації параметрів застосовувалась процедура ізотонічної розбивки [8]. Спочатку кожне значення ознаки замінюється на обчислене за формулою:

$$V_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad (6)$$

де x_{ij} — значення j -ї ознаки для i -го об'єкта.

Після цього кожному об'єктові ставиться у відповідність одне число, розраховане як:

$$w_i = \sum_{j=1}^m V_{ij} \quad (7)$$

Відстані між двома об'єктами визначають за формулою: $d_{ij} = |\omega_i - \omega_j|$. (8)

Таблиця 1

Показники масиву вхідних даних

Назва параметра	Характеристики параметра
1) Час спостереження	1999–2002 роки
2) Місце спостереження	Районні центри Черкаської області
3) Окисли азоту (NO), 4) сірчистий ангідрид (SO ₂), 5) окис вуглецю (CO), 6) вуглеводні (C _n (H ₂ O) _n , 7) діоксид марганцю (MnO ₂), 8) аміак (NH ₃), 9) сажа (SiO ₂), 10) толуол (C ₆ H ₅ CH ₃), 11) бутилацетат (CH ₃ (CH ₂) ₃ OOCH), 12) ацетон (CH ₃ COCH ₃), 13) легкі органічні сполуки (ЛОС), 14) оксид кальцію (CaO), 15) ксилол (C ₈ H ₁₀), 16) етилацетат (CH ₃ COOC ₂ H ₅), 17) тетраетилсвинець ((C ₂ H ₅) ₄ Pb), 18) бензин (C ₅ H ₁₂), 19) фенол (C ₆ H ₅ OH).	Концентрація речовин в повітрі жилої зони

За результатами кластеризації формувались додаткові показники шляхом перемноження значень існуючих показників, що ввійшли до одного кластеру. Якість моделей оцінювались за критерієм регулярності (8). В таблиці 2 подані результати досліджень.

Таблиця 2

Результати досліджень

№	Спосіб формування сигналів					
	Логарифмування добутоків		Зміна СКВ	Середнє арифметичне		Зміна СКВ
	Без кластеризації	З кластеризацією		Без кластеризації	З кластеризацією	
1	4,17	3,98	-4,59%	4,17	4,07	-2,52%
2	4,33	2,23	-48,53%	4,25	4,15	-2,25%
3	3,42	3,11	-9,09%	3,46	3,77	8,97%
4	2,55	2,27	-10,95%	2,54	2,87	12,99%

Як видно із таблиці для генерування несуміщених сигналів за результатами кластерного аналізу доцільно застосовувати процедуру логарифмування добутоків сигналів, які поєднуються в кластери. СКВ при цьому зменшується від 4,6% до 48,5%.

Висновки

Таким чином запропонований новий метод формування додаткових показників масиву вхідних даних для синтезу моделей об'єктів кризового моніторингу. Експериментально підтверджено ефективність застосування цього методу в алгоритмах індуктивного моделювання. Значення критерію регулярності результатів моделювання в досліджуваних умовах знижується в середньому на 18,3%.

Список літературних джерел

1. Голуб С.В. Методологія створення автоматизованих систем багаторівневого соціоекологічного моніторингу: автореф. дис. на здобуття наук. ступеня докт. техн. наук: спец. 05.13.06 «Інформаційні технології» / С.В. Голуб. – Київ. – 2009. – 36 с.
2. Индуктивный метод самоорганизации моделей сложных систем. Ивахненко А.Г. Киев: Наук. думка, 1981. 296 с.
3. Голуб С.В. Визначення структури інформаційних моделей в технологіях управління довкіллям // Вісник інженерної академії України. – 2007. – № 2. С. 35-42.
4. Голуб С.В. Визначення інформативності індуктивних моделей евристичних систем спостереження // Вісник Східноукраїнського національного університету імені Володимира Даля. - 2007. - № 5 (111), - Частина 1. - С. 265-269.
5. Ивахненко А.Г. Помехоустойчивость моделирования. А.Г. Ивахненко, В.С. Степашко. – Киев: Наук. думка, 1985. – 216 с.
6. Голуб С.В. Принцип проектування багаторівневих технологій інформаційного моделювання // Вісник інженерної академії України. – 2007. – № 1. – С. 28-34.
7. Голуб С.В. Підвищення інформативності первинного опису в технологіях індуктивного моделювання об'єктів вимірювання / С.В. Голуб // Вісник НТУУ „КПІ”. Серія приладобудування. – 2006. Вип.. 32. С. 118-122.
8. Лапач С.Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С.Н. Лапач, А.В. Чубенко, П.Н. Бабич. – К.: МОРИОН, 2001. – 408с.