

СТРУКТУРОВАНІ МОДЕЛІ ІНФОРМАЦІЙНИХ ПОЛІВ

1 Постановка задачі

Будь-яке інформаційне поле є складною динамічною системою (СДС) за визначенням. Задача ідентифікації СДС, як правило, описується як багатопільова задача оптимізації. У такій постановці вона формулюється в такий спосіб: необхідно знайти систему, з деякого припустимого класу систем, що є найкращою моделлю для заданої системи вхідних даних. Під терміном “найкраща” мається на увазі, що шукана модель здатна генерувати дані, найбільш близькі до заданих, чи, говорячи іншими словами, має найкраще наближення (найменша невідповідність) до даних. Модель також повинна бути найпростішою з множини доступних моделей. Крім критеріїв простоти і точності, можуть застосовуватися й інші критерії.

2 Вихідні дані

Для розробки методів і алгоритмів ідентифікації СДС був використаний підхід [1], при якому система розглядається як сукупність визначеного типу відносин, що дає можливість розробляти універсальні методи й алгоритми ідентифікації, застосовні до різних типів об'єктів. У рамках цього підходу виділяють наступні ієрархічні рівні представлення системи: моделі системи даних, генераційної системи, структурованої системи, метасистеми [1]. Задача ідентифікації вирішується на кожному з рівнів.

Вихідними даними для рішення задачі є:

- досліджувана система, задана у вигляді

$$X = \{X_i, i = \overline{1, n}\}, \quad (1)$$

де X - система з n нейтральних змінних (змінні не поділяються на вхідні і вихідні) з v_i рівнями квантування для кожної змінної, і послідовності N одночасних вимірів змінних X , представлених у вигляді матриці даних

$$D_{вх} = X \times N; \quad (2)$$

- клас систем, серед якого шукається рішення;
- критерії й обмеження.

Згідно [1] вихідна система може бути описана моделями на різних ієрархічних рівнях представлення системи.

© Т.А. Ліхоузова, О.С. Ліхоузов, 2006

3 Система даних

На рівні системи даних модель має вигляд:

$$D_x = (X, \vartheta, W, d), \quad (3)$$

де X – множина змінних, $\vartheta = \{\nu_i, i = \overline{1, n}\}$ – множина припустимих значень змінних, W – множина параметрів, у розглянутій задачі $W = \overline{1, N}$, і $d : W \rightarrow \vartheta$ – функція, що ставить у відповідність будь-якому значенню параметра один стан змінної. У залежності від задачі, що розглядається, функція d може бути визначена принаймні трьома способами. По-перше, вона може бути визначена користувачем, як це буває в задачах проектування систем. По-друге, її можна вивести із систем більш високих рівнів. По-третє, вона може бути результатом вимірів чи спостережень, як це буває при емпіричних дослідженнях. В останньому випадку модель системи даних приймає вигляд

$$D_x = (X, \vartheta, D_{ucx}). \quad (4)$$

У багатьох випадках для скорочення обчислювальних витрат на рішення задачі ідентифікації системи скорочують область припустимих значень змінних шляхом вибору іншого числа рівнів квантування змінних. У роботі [2] доведено, що найбільш ефективним з погляду реконструкційного аналізу є метод рівноімовірного квантування; там же запропонований алгоритм вибору оптимального числа рівнів квантування в залежності від кількості вхідних даних і обчислювальних обмежень. Тоді модель системи даних приймає вигляд:

$$D = (X, Q, D_{ucx}), \quad (5)$$

де $Q = \{q_i, i = \overline{1, n}\}$ – нова область припустимих значень змінних.

4 Генераційна система

На рівні генераційної системи модель має вигляд:

$$F = (D, M, f), \quad (6)$$

де M – маски, визначені на D , f – функція поводження, визначена через M і D [1]. За допомогою маски здійснюється перехід від вхідних змінних X до похідних змінних V . Для цілком упорядкованих параметричних множин маска може бути задана у вигляді вирізки з матриці, що представляє декартовий добуток

$$M_{\max} = X \times R, \quad (7)$$

де R – глибина маски. Послідовне переміщення маски по матриці даних дозволяє формувати коротез з

$$M \leq X \times R \quad (8)$$

змінних у кожному новому положенні.

У залежності від мети моделювання для синтезу моделі генераційної системи можна використовувати: метод [3], якщо в досліджуваній

системі нас цікавлять статичні взаємозв'язки між змінними; метод [4], якщо нас цікавлять динамічні взаємозв'язки між змінними; і метод [5], якщо нас цікавлять динамічні взаємозв'язки між змінними й обмеження дозволяють використовувати ту саму маску для всіх змінних. У [4] запропонований універсальний алгоритм реалізації всіх трьох методів.

5 Структурована система

Структуровані системи – це множини вихідних систем, систем даних чи генераційних систем. Вони потрібні для об'єднання декількох систем у великі. Для того, щоб таке об'єднання було коректним, необхідно, щоб окремі системи – елементи структурованої системи, були сумісні, тобто були одного типу і визначені на тій самій параметричній множині (умова сумісності) [1]. Крім того, ніякий елемент не повинний бути підсистемою іншого елемента тієї ж структурованої системи (вимога ненадмірності) [1]. Виконання цієї вимоги дозволяє уникнути перемішування рівнів окремих структурованих систем для того, щоб вони були ієрархічно упорядковані.

Для формального визначення моделі структурованої системи припустимо, що $S \subseteq S_1 \dots S_q$ нейтральна структурована система, що складається з $q = \overline{1, N_q}$ елементів (нейтральних систем того ж типу), що задовольняють вимогам сумісності і ненадмірності. Нехай, крім того,

$$V = \{v_i | i = \overline{1, n}\} \quad (9)$$

є множиною всіх змінних, що входять в елементи системи, і нехай qV - множина змінних елемента S_q . Тоді будь-який елемент однозначно ідентифікується множиною своїх змінних qV і

$$V = \bigcup_{q=1, N_q} {}^qV. \quad (10)$$

Різні типи структурованих систем позначаються стандартним для цього типу символом із префіксом S [1], що використовується як оператор і показує, що кілька систем визначеного типу об'єднано у велику систему. Так, позначення SD відповідає моделі структурованої системи, елементами якої є системи даних, а SF – моделі структурованої системи, елементами якої є генераційні системи.

$$SD = ({}^qV, {}^qD) \quad (11)$$

$$SF = ({}^qV, {}^qF) \quad (12)$$

Для синтезу моделі структурованої системи використовуємо метод [1], що ґрунтується на алгоритмах генерації структурних кандидатів і оцінці отриманих моделей структурованих систем за обраним критерієм.

Розглянемо нейтральну систему даних (1), (2) і модель генераційної системи, що представляє систему даних у термінах едальної маски (6). Кожна гіпотетична модель структурованої системи, що апроксимує систему

даних у рамках заданої маски, складається з: елементів; генераційних систем, зв'язаних з окремими елементами; зв'язків між елементами.

Якщо V позначає множину змінних, вхідних у задану модель генераційної системи, то кожен елемент розглянутої моделі структурованої системи визначається як підмножина V , тобто, якщо qV позначає множину змінних елемента S_q структури моделі, тоді ${}^qV \subseteq V$. Крім того, кожен елемент S_q отриманий шляхом проходу системи даних через єдину підмаску, визначену в термінах змінних, що входять у qV . При виконанні цих умов кожна модель структурованої системи може бути визначена єдиним способом у термінах сімейства підмножин V , де зв'язки між елементами визначені перетинанням пар підмножин V . Кожне сімейство підмножин V визначає єдину модель структурованої системи, сумісну з заданими даними і генераційними системами, але не усі вони придатні для розглянутої задачі. Значимі кандидати для гіпотетичних моделей структурованих систем повинні задовольняти наступним умовам [6].

Перше: множина змінних, вхідних у систему даних, еквівалентна множині змінних, що ідентифікує елементи моделі структурованої системи, тобто

$$\cup b = \bigcup_{{}^qV \in b} {}^qV = V, \quad (13)$$

де b позначає сімейство підмножин V , що єдиним способом визначає модель структурованої системи.

Друге: у моделі структурованої системи немає елементів, що ідентифіковані підмножиною змінних, котрі пов'язані через інші елементи тієї ж моделі.

Третє: у моделі структурованої системи немає елементів, що ідентифіковані порожньою множиною, тобто

$${}^qV \neq 0, \forall {}^qV \in b \quad (14)$$

Моделі структурованої системи, що задовольняють перерахованим умовам, є структурними кандидатами для заданої системи даних, у рамках заданої маски, а самі умови - це аксіоми структурних кандидатів, на основі яких можна побудувати алгоритм генерації решітки всіх можливих структурних кандидатів і визначити серед них найкращий [7].

Міри порівняння структурних кандидатів розглянуті в [1,7]. Найбільш універсальна серед них – узагальнена міра інформаційної відстані, розрахована по формулі

$$G({}^1f, {}^2f) = D\left({}^1f, \frac{{}^1f + {}^2f}{2}\right) + D\left({}^2f, \frac{{}^1f + {}^2f}{2}\right), \quad (15)$$

де 1f і 2f - довільні розподіли імовірностей, визначені на тій самій множині станів; $({}^1f + {}^2f)/2$ - розподіл імовірностей, отриманий узяттям середнього для кожної пари відповідних імовірностей з 1f і 2f ; D - інформаційна відстань, задана рівнянням

$$D(p_0, p_k) = \frac{1}{\log_2 |C|} \sum_{c \in C} p_0(c) \log_2 \frac{p_0(c)}{p_k(c)}, \quad (16)$$

де c - стан системи, C - множина усіх можливих станів системи, p_0 і p_k – розподіли імовірностей, визначені на тій самій множині станів, константа $1/\log_2 |C|$ – коефіцієнт, що забезпечує виконання співвідношення $0 \leq D(p_0, p_k) \leq 1$.

6 Метасистема

На рівні динамічної мережі (інакше – метасистеми [1]) для позначення використовуємо (подібно оператору S для структурованих систем) оператор M : поміщений перед позначенням системи визначеного типу, він позначає динамічну мережу, елементами якої є системи даного типу. Наприклад, MD - мережа, елементами якої є системи даних. Будь-яка динамічна мережа визначається як трійка

$$MY = (W, \Upsilon, r), \quad (17)$$

де $\Upsilon \in \{D, F, SD, SF\}$ – довільна множина систем, чиї параметричні множини є підмножинами W , r – процедура заміни, що реалізує визначену функцію заміни вигляду

$$r : W \rightarrow \Upsilon. \quad (18)$$

Сама задача синтезу динамічної мережі вирішується в три етапи:

1. визначаються ділянки стаціонарності (незмінності структури системи) на параметричній множині навчальної вибірки даних;
2. синтезуються моделі генераційних чи структурованих систем для кожної з ділянок стаціонарності;
3. синтезується процедура заміни моделі.

Якщо навчальна вибірка даних відсутня, то після виявлення кінця поточного ділянки стаціонарності просто набирається мінімальний обсяг даних з нової ділянки, необхідний для синтезу нової більш адекватної моделі.

На першому етапі можна використовувати різні методи виявлення розладки, наприклад, методи кумулятивних сум [8], усередненого відношення правдоподібності [9] чи метод ковзної вибірки спостережень [9]. Як випадкову послідовність для виявлення розладання можна використовувати як значення змінних з вихідної вибірки даних, так і визначені на них функції, наприклад, ентропію. Останнє дає можливість використовувати більш прості одномірні методи виявлення розладки. Для реалізації методу модифіковано один з відомих алгоритмів [10] для застосування до СДС.

З метою визначення якісних характеристик розглянутих методів і алгоритмів ідентифікації СДС було проведено ряд експериментів імітаційного моделювання [10]. Як досліджувана система використовувалася система з 4 змінних з 2 рівнями квантування і максимальною потужністю

маски 8 (приклад 4.3, с.202, [1]). Процедура реалізації одного експерименту така. По вхідній моделі генераційної системи ${}^T F$ генерується вибірка даних D обсягом 4000 значень. Потім вибирається найкраща маска і будуються моделі генераційної ${}^D F$ і структурованої ${}^S F$ систем. Для визначення, наскільки модель відповідає вихідній системі даних, виконується зворотний перехід від структурованої моделі до моделі генераційної системи ${}^S F$, по якій генерується вибірка даних для вихідних змінних. Рішення отримано за 8 ітерацій алгоритму з 15 можливих.

У результаті експериментів були отримані наступні залежності.

Залежність інформаційної відстані (розрахованого по формулі (15)) від числа спостережень для вдалих і невдалих структурних кандидатів на одному рівні уточнення (удалий, найкращий серед невдалих, середнє по невдалим) представлена на рис. 1. Незважаючи на те, що на вигляд кривих впливає число змінних і потужність множин їхніх станів, а також використовуваний тип відстані, з ростом числа спостережень ці відстані завжди зменшуються, а відстань для коректної структури прагне до нуля.

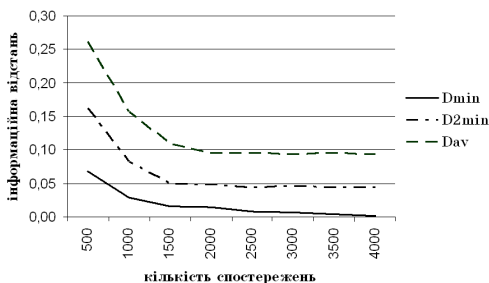


Рис. 1 – Залежність інформаційної відстані від числа спостережень.

На рис. 2 показаний взаємозв'язок множин станів з ненульовими імовірностями для трьох генераційних систем ${}^T F$, ${}^D F$ і ${}^S F$, що беруть участь в обчислювальному експерименті; позначимо ці множини відповідно ${}^T X$, ${}^D X$ і ${}^S X$. На нижньому графіку показана частка тих станів ${}^T F$, що мають в ${}^D F$ (це відбувається через недолік даних), тобто $({}^D X / {}^T X) \cdot 100$; на верхньому графіку представлений відсоток станів ${}^T F$, що мають в ${}^S F$, тобто $({}^S X / {}^T X) \cdot 100$. З цих графіків ясно видно, що ${}^D X \subseteq {}^S X \subseteq {}^T X$.

На графіках рис. 3 порівнюються інформаційні відстані між справжньою системою ${}^T F$ і даними, отриманими по моделях рівня генераційної ${}^D F$ і структурованої ${}^S F$ систем, у залежності від числа спостережень. Реконструйована система (тобто модель рівня структурованої системи) ближче до справжньої системи, ніж модель генераційної системи. Це зумовлено тим, що при недоліку даних коректна модель структурованої системи дозволяє відновити деякі стани системи, що не ввійшли у вибірку даних.

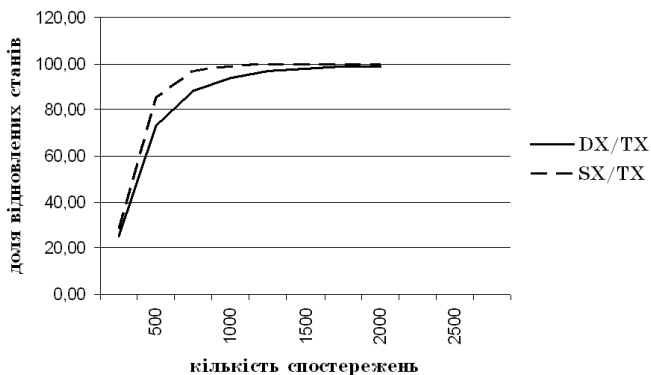


Рис. 2 – Взаємозв’язок множин станів з ненульовими імовірностями.

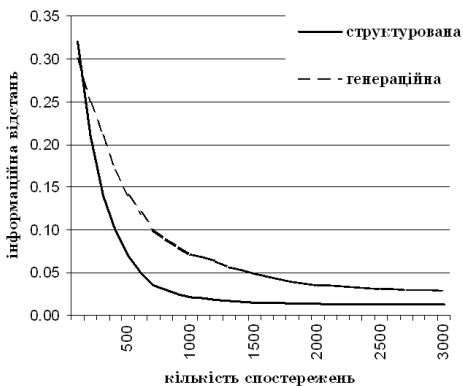


Рис. 3 – Порівняння інформаційних відстаней між справжньою системою $T F$ і даними, отриманими по моделях.

Засоби ідентифікації і моделювання складних динамічних систем, реалізовані у вигляді програмного комплексу, дозволяють створювати на сумісній техніці дослідницькі і промислові системи для ГВС, призначених для роботи без участі людини, наприклад, у задачах екстремальної екології.

Література

1. Клір, Дж. Системологія: автоматизація рішення системних задач/ М.: Радіо і зв'язок, 1990 – 539с.
2. Литвинов Є. М., Ліхоузов О. С., Ліхоузова Т. А. Модель системи даних в реконструкційному аналізі складних динамічних систем // Системні технології: комп'ютерна обробка експериментальних даних, 3, Дніпропетрівськ, 1998 – с.132-139
3. Conant, R. Detection and analysis of dependency structures// International Journal of General Systems, 7, 1981, pp.81-91
4. Литвинов Є. М., Ліхоузов О. С., Ліхоузова Т. А. Алгоритм ідентифікації складних динамічних систем// “Проблеми створення нових машин і технологій” Збірник наукових праць КГПІ, 1, Кременчук, 1999 – с.217-222
5. Klir, G.J. On the representation of activity arrays// International Journal of General Systems, 2, N3, 1975, pp.149-168
6. Klir G.J. and H.J.J.Uyttenhove, Computerized Methodology for Structure Modeling// International Journal of General Systems, Vol. 3, No. 3, 1976, pp.29-66
7. Литвинов Є. М., Ліхоузов О. С., Ліхоузова Т. А. Алгоритм структурної ідентифікації складних динамічних систем// Системні технології: комп'ютерна обробка експериментальних даних, 3, Дніпропетрівськ, 1998 – с.143-151
8. Бродський Б.Е. Асимптотично оптимальні методи в задачі якнайшвидшого виявлення розладки// АіТ, 1995, 9, с.50-59
9. Грушецька М.И., Кописницький Т.Н. Порівняння ефективності методів ковзного іспиту і контрольної вибірки при структурній ідентифікації// АіТ, 1996, 9, с.43-50
10. Литвинов Є. М., Ліхоузов О. С., Ліхоузова Т. А. Ідентифікація складних динамічних систем за допомогою Д-мережі // Тези конференції “Dynamical Systems Modeling and Stability Investigation” Київ, 2001 – с.186