

УДК 62.50

НЕЧІТКІ ЛОГІЧНІ СИСТЕМИ З ВИКОРИСТАННЯМ НЕЧІТКИХ МНОЖИН ВИЩИХ ТИПІВ

Н.Р. КОНДРАТЕНКО, С.М. КУЗЕМКО

Запропоновано показник якості функціонування та метод побудови нечітких логічних систем із використанням нечітких множин вищих типів на основі експериментальних даних. Проведено дослідження впливу кількості пропусків у вхідних даних на можливість отримання вихідного значення.

ВСТУП

Експериментальні дані, що характеризують певну предметну область, як правило, містять різноманітні невизначеності. Вхідні значення та дані для настройки параметрів нечіткої моделі у багатьох випадках містять шуми або пропуски. Також можлива неузгодженість думок експертів, розбіжність визначення лінгвістичних термів різними людьми і т. ін. [1–3].

Сучасні підходи до розв'язання задачі опису різних невизначеностей (у тому числі і наявності пропусків як в експериментальних, так і у вхідних даних) пов'язують з інтервальними функціями належності [1]. У роботі [2] запропоновано метод побудови нечітких моделей на основі експериментальних даних, що містять пропуски, та показник якості їх функціонування з використанням інтервальних функцій належності. При їх використанні в якості вихідної величини нечіткої логічної системи виступає інтервал значень, що накладає обмеження на можливості опису різноманітних невизначеностей, наявних у предметній області.

ПОСТАНОВКА ЗАДАЧІ

Поставимо задачу розробки методу побудови нечітких моделей на основі експериментальних даних, що містять пропуски, показника якості функціонування нечітких моделей з використанням нечітких множин вищих типів та дослідження впливу кількості пропусків у вхідних даних на можливість отримання вихідного значення.

Побудова нечітких моделей із експериментальних даних, які містять пропуски, з використанням нечітких множин вищих типів

Нехай є експериментальна вибірка X .

$$X = \{X_1, X_2, \dots, X_n\},$$

де $X_i = (x_{1i}, x_{2i}, \dots, x_{ki}, y_i)$, $i = 1, \dots, n$; n — кількість експериментальних прикладів; k — кількість вхідних змінних; y — вихідна величина.

В експериментальній вибірці значення деяких x_{ij} невідомі, тобто існує певна множина $H \subset [(1 \dots k) \times (1 \dots n)]$ така, що $x_{ij \in H} = \emptyset$.

Базу правил можна визначити таким чином:

$$R_1 : \text{IF } x_1 = {}^{(q)}\tilde{A}_{11 \notin H} \text{ AND } \dots x_i = {}^{(q)}\tilde{A}_{1i \notin H} \text{ AND } \dots \text{ AND } x_k = {}^{(q)}\tilde{A}_{k1 \notin H} \text{ THEN } y = {}^{(p)}\tilde{B}_1,$$

$$R_n : \text{IF } x_1 = {}^{(q)}\tilde{A}_{1n \notin H} \text{ AND } \dots \text{ AND } x_i = {}^{(q)}\tilde{A}_{1i \notin H} \dots \text{ AND } x_k = {}^{(q)}\tilde{A}_{kn \notin H} \text{ THEN } y = {}^{(p)}\tilde{B}_m,$$

де ${}^{(q)}\tilde{A}_{ij}$, $i = 1, \dots, n$, $j = 1, \dots, k$ — нечітка множина типу q ; ${}^{(p)}\tilde{B}_i$, $i = 1, \dots, n$ — нечітка множина типу p .

Центри нечітких множин визначаються відповідними експериментальними даними [1, 2].

Алгоритм прийняття рішень при фіксованому вхідному векторі $X^* = \{X_1^*, X_2^*, \dots, X_n^*\}$, який не містить пропусків, має вигляд

$$y = \bigvee_{j=1}^m \bigwedge_{i=1}^n \mu^{(q)\tilde{A}_{ij \notin H}}(x_j^*, \mu_1, \dots, \mu_{q-1}),$$

де $\mu^{(q)\tilde{A}_{ij}}(x_j, \mu_1, \dots, \mu_{q-1})$ — функція належності параметра x_j нечіткому терму ${}^{(q)}\tilde{A}_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n$; \vee — операція АБО, в якості якої найбільш часто використовується максимум; \wedge — операція І, в якості якої найбільш часто використовується мінімум.

Для прийняття рішення при пропусках у вхідних даних будемо використовувати алгоритм, запропонований у роботі [2].

- Є вхідний вектор параметрів, що містить пропуски.
- З початкової експериментальної вибірки виключаються стовпці, для яких невідомі вхідні значення, та формується нова експериментальна вибірка.
- Проводиться генерування нечіткої моделі на основі скороченої експериментальної вибірки.
- Визначаються вихідні параметри.

Показник якості функціонування нечітких моделей

Будемо вважати відомим:

- 1) множину вхідних параметрів $X = \{x_1, x_2, \dots, x_n\}$;
- 2) вихідні параметри y ;
- 3) систему правил, яка описує певну предметну область (для спрощення будемо вважати, що в ній немає пропусків)

$$R_1 : \text{IF } x_1 = {}^{(q)}\tilde{A}_{11} \text{ AND } \dots \text{ AND } x_n = {}^{(q)}\tilde{A}_{1n} \text{ THEN } y = {}^{(p)}\tilde{B}_1,$$

$$R_m : \text{IF } x_1 = {}^{(q)}\tilde{A}_{m1} \text{ AND } \dots \text{ AND } x_n = {}^{(q)}\tilde{A}_{mn} \text{ THEN } y = {}^{(p)}\tilde{B}_m,$$

де n — кількість змінних; m — кількість правил; ${}^{(q)}\tilde{A}_{ij}$, $i=1, \dots, m$, $j=1, \dots, n$ — нечітка множина типу q ; ${}^{(p)}\tilde{B}_i$, $i=1, \dots, m$ — нечітка множина типу p ;

4) механізм логічного висновку [2]

$$y = \bigvee_{j=1}^m \bigwedge_{i=1}^n \mu^{(q)}\tilde{A}_{ij} (x_j, \mu_1, \dots, \mu_{q-1}).$$

При використанні інтервальних функцій належності значення показника якості функціонування нечітких моделей дорівнює [2]

$$UN = \frac{\sum_{i=1}^n \sum_{j=1}^m \left(\int_{\underline{x}_i}^{\overline{x}_i} \mu^{(2)}\tilde{A}_{ij} dx_i - \int_{\underline{x}_i}^{\overline{x}_i} \underline{\mu}^{(2)}\tilde{A}_{ij} dx_i \right)}{nm},$$

де $\overline{\mu}^{\tilde{A}_{ij}}$ та $\underline{\mu}^{\tilde{A}_{ij}}$ — верхня та нижня границі інтервальної функції належності $\mu^{\tilde{A}_{ij}}$; \underline{x}_i та \overline{x}_i — максимальне і мінімальне значення змінної x_i , де n — кількість змінних; m — кількість правил; UN — показник якості функціонування нечітких моделей.

Запропонуємо показник якості функціонування нечітких моделей для нечітких множин типу 2 (рис. 1), розглянутих у роботі [1].

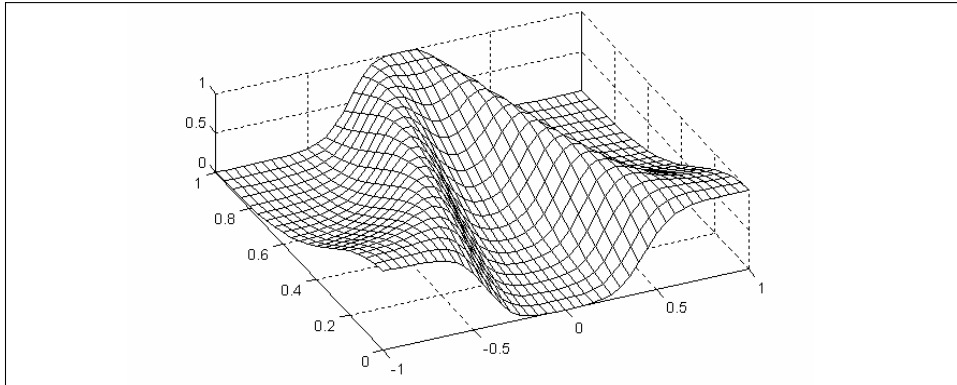


Рис. 1. Приклад функції належності нечіткої множини типу 2

Кожному значенню вхідної змінної ставиться у відповідність нечітка множина типу 1 [1]. Множина функцій належності нечітких множин типу 1 утворює поверхню, яка в певному сенсі визначає невизначеність опису нечіткої множини типу 2. Якщо для інтервальних функцій належності показник якості функціонування нечітких моделей визначається площею області невизначеності, то для тривимірних функцій належності нечітких множин типу 2 цілком природно використовувати поняття об'єму. Отже

$$UN = \frac{\sum_{i=1}^n \sum_{j=1}^m \left(\int_{x_i}^{\bar{x}_i} \int_0^1 \mu^{(2)\tilde{A}_{ij}}(x_i, \mu_1) dx_i d\mu_1 \right)}{nm}.$$

$UN \in [0,1]$, причому $UN = 0$ відповідає повній відсутності невизначеності, а $UN = 1$ — повній невизначеності.

Узагальнимо результати, отримані для інтервальних функцій належності та функцій належності нечітких множин типу 2 на випадок нечітких множин загального типу.

$$UN = \frac{\sum_{i=1}^n \sum_{j=1}^m \left(\int_{x_i}^{\bar{x}_i} \int_0^1 \dots \int_0^1 \mu^{(q)\tilde{A}_{ij}}(x_i, \mu_1, \dots, \mu_{q-1}) dx_i d\mu_1, \dots, d\mu_{q-1} \right)}{nm}.$$

За допомогою розробленого показника можна порівняти множину побудованих нечітких моделей з використанням нечітких множин вищих типів. Найкращою моделлю буде та, яка характеризується найменшим значенням показника та функціями належності з максимальною невизначеністю при умові адекватності відображення експериментальних даних.

Комп'ютерний експеримент

Розглянемо вплив кількості пропусків у вхідних даних на можливість отримання прийняттого вихідного значення на прикладі задачі діагностування гіпотиреозу. Предметна область діагностування гіпотиреозу має такі характеристики [4]:

- 44 вхідних параметри.
- Одна вихідна величина (відсутність або три ступеня важкості захворювання).
- 387 експериментальних даних про предметну область, що містять пропуски. Загальна кількість пропущених значень — 1205.

Модель діагностування гіпотиреозу побудуємо на інтервальних функціях належності, для опису яких оберемо модифіковану гауссову форму.

Загальний вигляд інтервальної функції належності

$$\mu(x) = e^{-\left(\frac{x-b}{[\min(c), \max(c)]}\right)^2},$$

де $[\min(c), \max(c)]$ — діапазон зміни параметру c гауссової функції належності.

Для визначення діапазону зміни параметру c використаємо алгоритм, який містить дві частини: 1) побудова звичайної нечіткої моделі по зменше-

ній вибірці і 2) знаходження діапазонів зміни параметру c функцій належності шляхом максимального розширення діапазонів зміни при збереженні адекватності на тестових даних [2].

Дослідження впливу кількості пропусків на функціональні можливості нечіткої моделі почнемо з граничних випадків: невідома лише одна вхідна величина та відоме одне вхідне значення.

Розглянемо випадок відсутності значення одного з параметрів. Кількість можливих варіантів — 43. Побудуємо 43 моделі та проаналізуємо, як відсутність значення одного з параметрів впливає на функціональні можливості моделі (кількість отриманих правил, значення помилки на навчальних, тестових та контрольних даних, а також значення показника якості функціонування моделі). Проведені дослідження дозволяють визначити вплив кожного з параметрів на вихідну величину та скоротити загальну кількість вхідних параметрів.

Обробка результатів експериментів показала, що виключення будь-якого параметру фактично не вплинуло на функціональні можливості моделі. Для всіх 43 побудованих моделей кількість правил становила 132, всі моделі безпомилково функціонують на навчальних, тестових та контрольних даних. Єдина відмінність від моделі, яка враховує 43 параметри, — незначне збільшення значення показника якості функціонування моделі на величину 0,0001 для всіх 43 моделей. Отриманий результат показує, що незнання значення одного з вхідних параметрів фактично не впливає на неможливість отримання задовільного вихідного значення, а також що жоден з параметрів не має суттєвого впливу на визначення кінцевого діагнозу. Даний результат був підтверджений експертом.

Розглянемо інший граничний випадок: відоме значення одного параметра, а значення інших 42 невідомі. Побудуємо 43 моделі, кожна з яких буде містити по одному вхідному та одному вихідному параметру.

Обробка результатів експерименту показує, що всі побудовані моделі містять по одному правилу, яке ставить у відповідність будь-якому вхідному значенню весь можливий діапазон вихідної величини (рис 2).

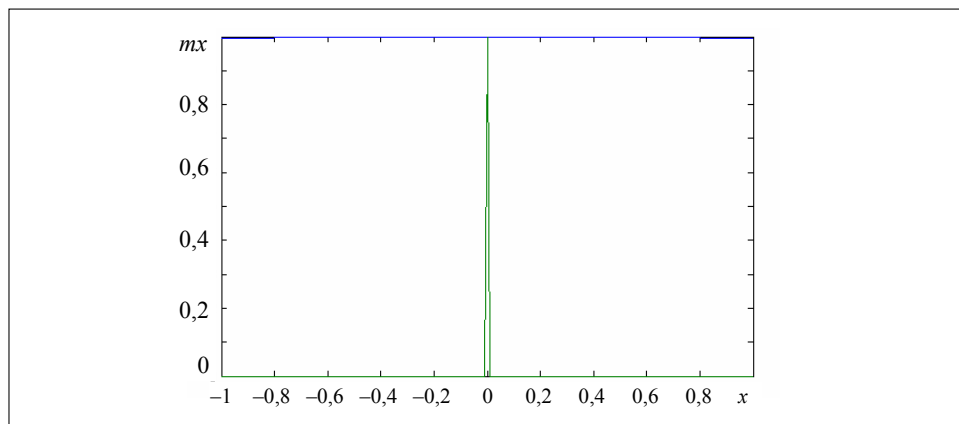


Рис. 2. Приклад отриманої інтервальної функції належності

Значення показника якості функціонування для всіх моделей дорівнює 0,9999. На навчальних, тестових та контрольних даних всі моделі в якості вихідної величини показали весь можливий діапазон діагнозів. Таке функціонування моделей відповідає обережній поведінці (наявної інформації об'єктивно було недостатньо для визначення вихідної величини). Також зробимо висновок, що лише за одним параметром неможливо визначити діагноз, тобто жоден з параметрів не є визначальним.

Проведемо дослідження впливу кількості пропусків на кількість отриманих правил та значення показника якості функціонування моделі (рис. 3, 4).

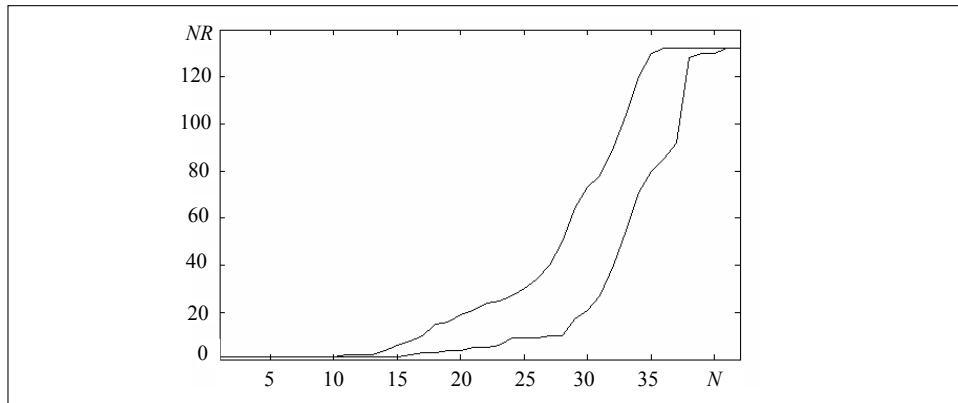


Рис. 3. Залежність кількості правил від кількості відомих значень параметрів

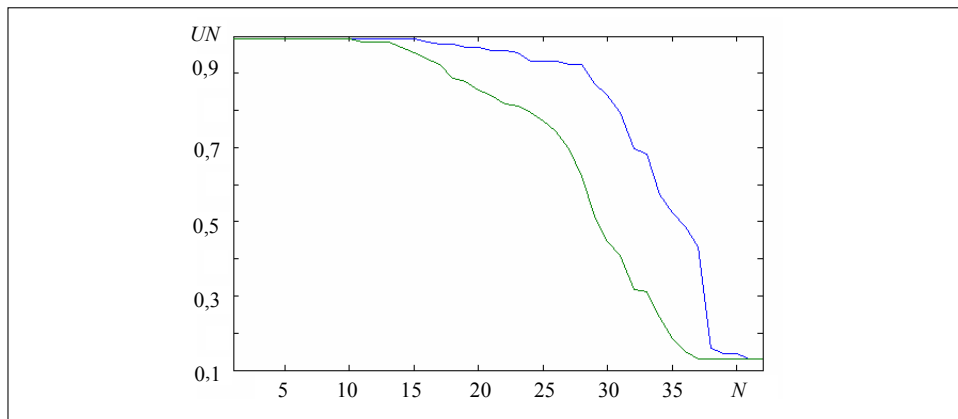


Рис. 4. Залежність значення показника якості функціонування нечіткої моделі від кількості відомих значень параметрів

Експерименти показали, що всі моделі функціонують таким чином: вихідне значення виступає або як діагноз, що співпадає з верифікованим, або як множина діагнозів, що його включає. Причому це відбувається на всіх вибірках (навчальній, тестовій та контрольній). Тобто можна стверджувати: моделі обережно визначають кінцевий діагноз на основі наявної інформації, якої може бути об'єктивно недостатньо для точного визначення вихідної величини (рис. 5–7).

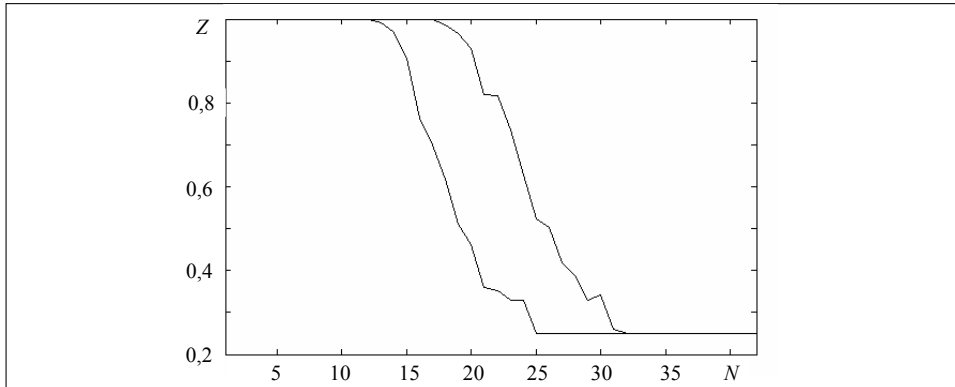


Рис. 5. Залежність середнього розкиду вихідної величини від кількості відомих значень параметрів на навчальних вибірках

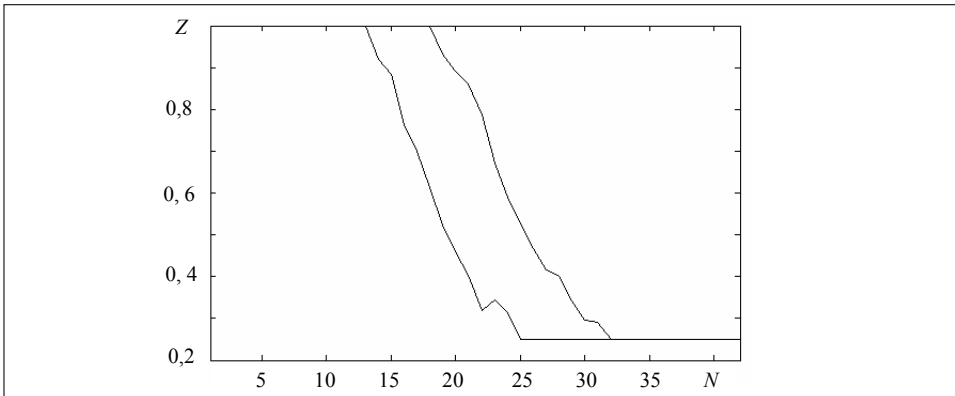


Рис. 6. Залежність середнього розкиду вихідної величини від кількості відомих значень параметрів на тестових вибірках

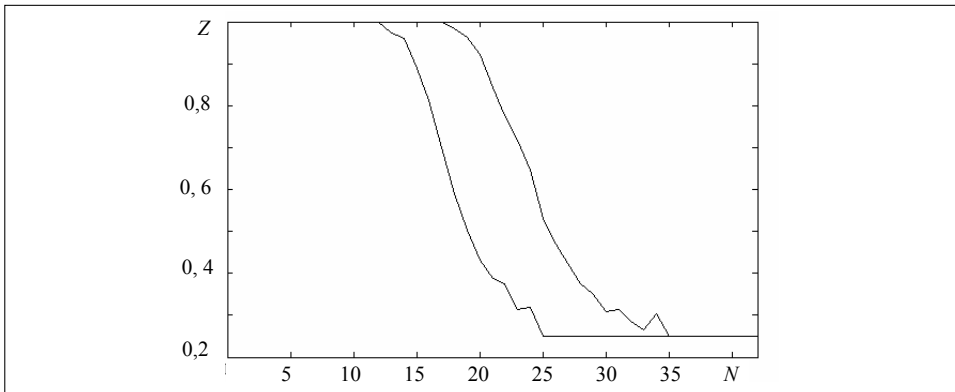


Рис. 7. Залежність середнього розкиду вихідної величини від кількості відомих значень параметрів на контрольних вибірках

Параметр Z визначає середній розкид вихідної величини і обчислюється за формулою

$$Z = \frac{\sum_{i=1}^n k_i}{4n},$$

де n — кількість експериментів; k_i — кількість вихідних діагнозів, які визначила модель для i -го експерименту.

Аналізуючи отримані результати, будемо стверджувати, що зменшення кількості відомих значень параметрів дає збільшення невизначеності нечіткої моделі та зменшення кількості правил. При 32...42 відомих значеннях параметрів можна точно визначити вихідну величину, яка в переважній більшості випадків не має невизначеності — це буде один, конкретний діагноз. При 1...13 відомих значеннях параметрів вихідну величину визначити взагалі неможливо (підтверджується експертом). При 14...31 відомому значенні параметрів як вихідна величина (у залежності від наявної інформації) виступає або точне значення, або множина діагнозів (пара сусідніх ступенів важкості, інформація про наявність захворювання або весь діапазон значень). У процесі функціонування моделі, якщо вихідне значення містить невизначеність, додавання додаткових значень зменшує загальну невизначеність моделі.

ВИСНОВКИ

Отримані теоретичні результати дозволяють реалізувати метод побудови нечітких логічних систем на основі експериментальних даних, що містять пропуски, та ввести показник якості функціонування нечітких моделей з використанням нечітких множин вищих типів. Експериментальні дослідження продемонстрували можливість отримання вихідного значення при пропусках у вхідних даних.

Подальші дослідження пов'язуються з розробкою методів побудови оптимальних нечітких логічних систем в умовах невизначеності.

ЛІТЕРАТУРА

1. *Mendel J.M.* Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. — Upper Saddle River: Prentice-Hall, 2001. — 450 p.
2. *Кондратенко Н.Р., Зелінська Н.Б., Куземко С.М.* Нечіткі логічні системи з врахуванням пропусків в експериментальних даних // *Наук. вісті НТУУ «КПІ»*. — 2004. — № 5. — С. 37–41.
3. *Цельх А.Н., Тимошенко Р.П.* Некоторые теоретико-множественные операции над интервальными нечеткими множествами в моделях искусственного интеллекта // *Перспективные информационные технологии и интеллектуальные системы*. — 2001. — № 2. — С. 69–76.
4. *Кондратенко Н.Р., Зелінська Н.Б., Куземко С.М.* Діагностика гіпотиреозу на основі нечіткої логіки з використанням інтервальних функцій належності // *Наук. вісті «КПІ»*. — 2003. — № 4. — С. 52–58.

Надійшла 15.12.2005