

Штовба С.Д., Мазуренко В.В., Панкевич О.Д. Вплив кількості нечітких правил на точність бази знань Мамдані // Вісник Хмельницького національного університету. Технічні науки. – 2011. – №2. – С. 185–188.

УДК 658.012

С.Д. ШТОВБА, В.В. МАЗУРЕНКО, О.Д. ПАНКЕВИЧ

Вінницький національний технічний університет

## ВПЛИВ КІЛЬКОСТІ НЕЧІТКИХ ПРАВИЛ НА ТОЧНІСТЬ БАЗИ ЗНАНЬ МАМДАНІ

*Досліджується задача ідентифікації багатофакторних залежностей за допомогою нечітких баз знань типу Мамдані. Проведені комп'ютерні експерименти засвідчили квадратичну залежність точності нечіткої бази знань Мамдані від кількості правил.*

*The work is devoted to the multifactor dependence identification problem with Mamdani-type fuzzy knowledge bases. Carried computational experiments show that the dependence between Mamdani-type knowledge base accuracy and number of the rules is quadratic.*

Ключові слова: ідентифікація, нечітка бази знань, повнота бази знань, точність.

### Вступ

Сьогодні все частіше моделювання багатофакторних залежностей в техніці, економіці, медицині, соціології, будівництві, сільському господарстві, спорті та в інших областях здійснюють за допомогою нечітких баз знань. Нечіткою базою знань називається сокупність нечітких правил “Якщо – тоді”, яка задає взаємозв'язок між входами та виходами досліджуваного об'єкту. В монографіях з проектування нечітких систем, наприклад в [1–4], представлені різноманітні підходи до вирішення прикладних задач за допомогою різних моделей баз знань, але в них немає відповіді на важливе практичне питання про необхідну кількість правил в базі знань. В нечіткому моделюванні найчастіше використовують базу знань Мамдані, в якій антецеденти та консеквенти описано нечіткими множинами. Тому, метою дослідження є виявлення впливу кількості правил на якість нечіткої бази знань Мамдані. Виявлення такої залежності дозволить визначити мінімальну кількість нечітких правил для адекватного опису досліджуваної залежності. В подальшому отриману у такий спосіб компактну нечітку базу знань можна краще і швидше навчити за експериментальними даними.

### Постановка задачі

Вважатимемо відомою вибірку з  $M$  пар експериментальних даних, що пов'язують фактори впливу  $X = (x_1, x_2, \dots, x_n)$  з виходом  $y$  досліджуваної залежності:

$$(X_r, y_r), \quad r = \overline{1, M}, \quad (1)$$

де  $X_r$  – вхідний вектор в  $r$ -ому рядку вибірки;  $y_r$  – відповідне вихідне значення.

Позначимо через  $y = F(N, X)$  модель на основі нечіткої бази знань Мамдані з  $N$  правилами, що пов'язують входи  $X$  з виходом  $y$  досліджуваної залежності. Точність нечіткої моделі визначимо через середню квадратичну нев'язку на вибірці (1):

$$RMSE = \sqrt{\frac{1}{M} \sum_{r=1, M} (y_r - F(N, X_r))^2} . \quad (2)$$

За фіксованого нечіткого розбиття вхідних та вихідної змінних можна згенерувати кілька нечітких баз знань Мамдані з одним і тим самим числом правил ( $N$ ). Тому задачу дослідження поставимо, як знаходження залежності точності  $RMSE$  від обсягу  $N$  бази знань для найкращого, найгіршого та середнього випадків. Побудову таких кривих навчання здійснимо експериментально на двох еталонних залежностях з подальшою апроксимацією аналітичними моделями.

### Нечітке виведення за базою знань Мамдані

Нечітку базу знань Мамдані запишемо таким чином [1]:

$$(x_1 = \tilde{a}_{1j} \text{ та } x_2 = \tilde{a}_{2j} \text{ та } \dots \text{ та } x_n = \tilde{a}_{nj}) \Rightarrow y = \tilde{d}_j, \quad j = \overline{1, N}, \quad (3)$$

де  $\tilde{a}_{ij}$  та  $\tilde{d}_j$  – нечіткі множини термів вхідних та вихідної змінної з функціями належності  $\mu_j(x_i)$  та  $\mu_{d_j}(y)$ ,  $i = \overline{1, n}$ ,  $j = \overline{1, N}$ .

Ступінь виконання антецедента  $j$ -го правила для поточного вхідного вектора  $X^* = (x_1^*, x_2^*, \dots, x_n^*)$  розраховують так:  $\mu_j(X^*) = \min(\mu_j(x_1^*), \mu_j(x_2^*), \dots, \mu_j(x_n^*))$ ,  $j = \overline{1, N}$ . Результатом виведення за  $j$ -им правилом бази знань буде нечітка множина

$$\tilde{d}_j^* = \text{imp}(\tilde{d}_j, \mu_j(X^*)), \quad j = \overline{1, N}, \quad (4)$$

де  $\text{imp}$  позначає імплікацію, яку реалізують операцією мінімуму.

Результат виведення за усіма правилами знаходять агрегуванням нечітких множин (4)  $\tilde{y}^* = \text{agg}(\tilde{d}_1^*, \tilde{d}_2^*, \dots, \tilde{d}_N^*)$ , знаходячи максимум функцій належності. Чітке значення вихода  $y^*$ , яке відповідає вхідному вектору  $X^*$ , визначається дефаззифікацією нечіткої множини  $\tilde{y}^*$  за методом центра тяжіння.

### Комп'ютерні експерименти

Експерименти проведемо для 2 еталонних залежностей (рис. 1) – неспадної та унімодальної

$$y = x_1 \sqrt{x_2}, \quad a \in [2; 22], b \in [2; 14], \quad (5)$$

$$y = -x_1^2 - x_2^2, \quad a \in [-7; 3], b \in [-5; 5], \quad (6)$$

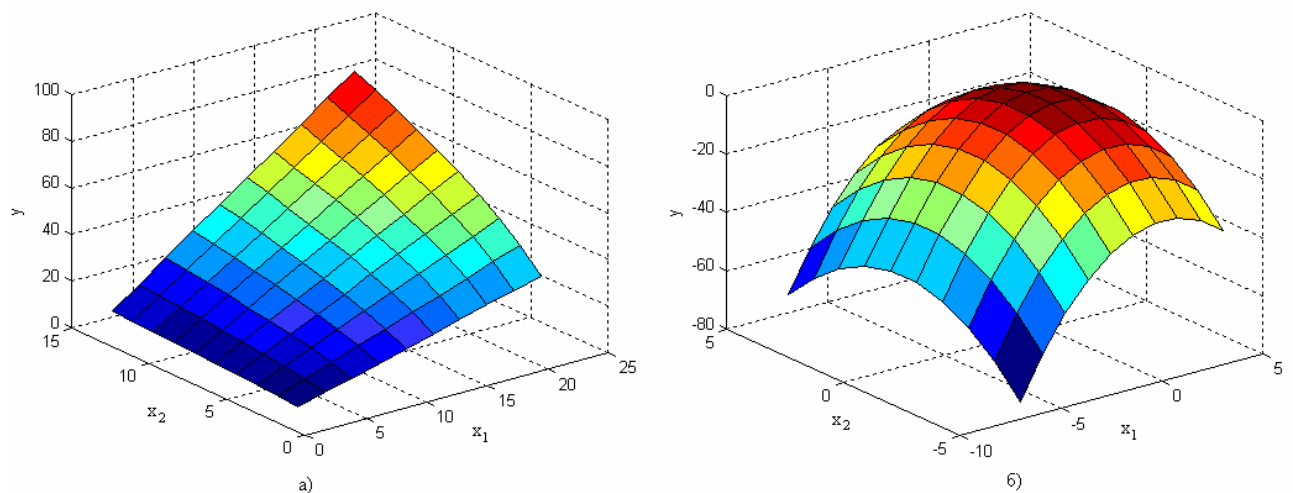


Рис. 1. Графіки еталонних залежностей  
 а) неспадна залежність (5), б) мультимодальна залежність (6)

Для кожної нечіткого розбиття експерименти проводилися за такою схемою:

- 1) згенерувати тестову вибірку з 100 точок;
- 2) згенерувати повний список з  $N_{\max}$  адекватних нечітких правил;
- 3) синтезувати усі можливі нечіткі бази з  $N$  правил,  $N = \overline{1, N_{\max}}$  ;
- 4) для кожної нечіткої бази знань розрахувати нев'язку на тестовій вибірці за формулою (2);
- 5) для кожної множини нечітких баз знань одного розміру знайти мінімальну  $RMSE_{\min}$  ,  
 максимальну  $RMSE_{\max}$  та середню  $RMSE_{mean}$  ;
- 6) побудувати графіки залежностей  $RMSE_{\min}$  ,  $RMSE_{mean}$  та  $RMSE_{\max}$  від  $N$  .

Вихідна змінна оцінена 5-ма термами. Для неспадної еталонної залежності для оцінки вхідних змінних використовувалося 2, 3 та 4 термів, тобто експерименти проведено для таких 9-ти нечітких розбиттів вхідних змінних: 2x2, 2x3, 2x4, 3x2, 3x3, 3x4, 4x2, 4x3 та 4x4. Відповідно, максимальна кількість адекватних нечітких правил ( $N_{\max}$ ) склала 4, 6, 8, 6, 9, 12, 8, 12 та 16. Для унімодалної використовувалося таке нечітких розбиття вхідних змінних: 3x3, 3x4, 3x5, 4x3, 4x4 та 5x3. Консеквентом кожного із цих правил обирався терм вихідної змінної з максимальним ступенем належності для значення еталонної функції для ядер нечітких термів антецедента. Нечітке розбиття здійснено за допомогою гаусових функцій належності, ядра яких рівномірно розподілено на діапазоні вхідних та вихідної змінних. Коefіцієнт концентрації функцій належності прийнято рівним  $c = \Delta core / 2.4$ , де  $\Delta core$  – відстань між ядрами сусідніх термів. Для уникнення ефекту звуження діапазону вихідних значень розширено носій нечітких множин вихідної змінної згідно до [5].

На протязі одного обчислювального експерименту перевірялось від  $2^{2 \cdot 2} - 1 = 15$  до  $2^{4 \cdot 4} - 1 = 65535$  нечітких баз знань, відповідно здійснено від 1500 до 6553500 нечітких виведень. Всього проведено  $2 \cdot 9 = 18$  експериментів – 9 для кожної із 2 еталонних залежностей. Результати експериментів (рис. 2) показали, що в середньому нев'язка  $RMSE_{mean}$  спадає зі збільшенням кількості нечітких правил і досягає мінімуму за повної бази знань. Якщо вдало підібрати комбінацію правил, тоді нев'язка  $RMSE_{\min}$  стає суттєво меншою і її мінімум досягається за неповної бази знань. Для найкращого випадку на кривих навчання добре простежується «плато насичення», коли додавання нових правил майже не змінює адекватність нечіткої моделі [6]. Бази знань з цього «плато насичення» назвемо прийнятними. Планку для них призначимо у вигляді 50% перевищення нев'язки у порівнянні з найкращою нечіткою моделі.

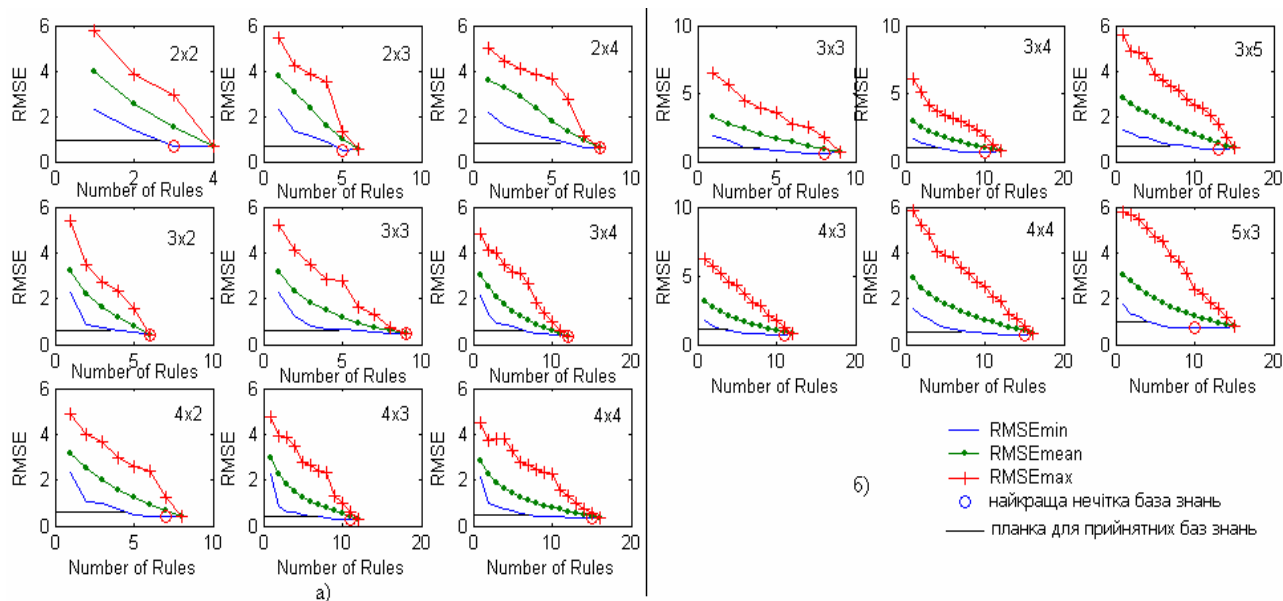


Рис. 2. Криві навчання нечітких баз знань Мамдані а) для залежності (5); б) для залежності (6)

Проведені квадратичні апроксимації експериментальних даних (рис. 3) свідчать, що повну нечітку базу знань Мамдані можна скоротити в 2–3 рази без великих втрат точності. Базы знань, які містять біля 80% від максимальної кількості правил, мають найвищу точність.

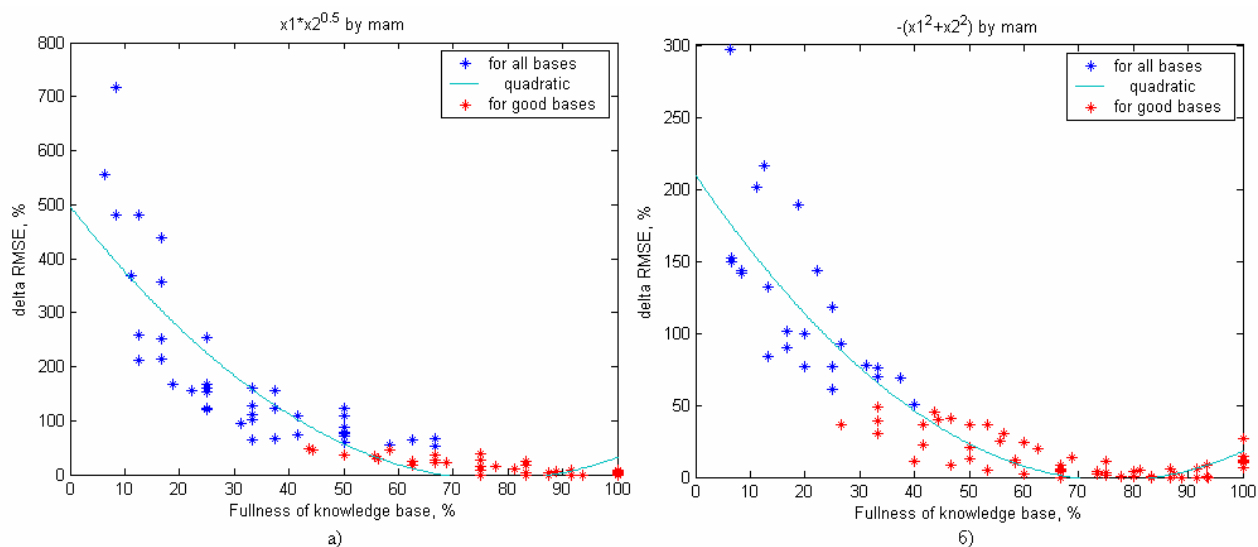


Рис. 3. Залежність збільшення похибки від повноти бази знань Мамдані - а) для залежності (5); б) для залежності (6)

### Висновки та подальші дослідження

Проведені комп'ютерні експерименти показують, що повна нечітка база знань Мамдані не дає найменшу похибку ідентифікації. На нашу думку, це обумовлено деякими суперечливостями на границях нечіткого розбиття через взаємодію великої кількості правил. Найменша помилка найчастіше спостерігається для баз знань, що наповненні правилами на 80-85%. Встановлено, що залежність точності нечіткої бази знань від її розмірності є квадратичною. Якщо вдало скоротити повну бази знань втричі, тоді помилка ідентифікації збільшиться лише на 30-50% в порівнянні з найкращим варіантом. Такі компактні база знань є прозорішими та легше навчаються через меншу складність відповідної задачі оптимізації. Під час створення нової нечіткої моделі дослідивши всього декілька наборів правил, наприклад, найменший, з 80% заповненням та повний,

можна отримати квадратичну апроксимацію точності від повноти бази знань.

Подальші дослідження будуть спрямувати на підтвердження отриманих експериментальних висновків на інших типах залежності та для інших форматів нечітких баз знань. Крім того, варто дослідити вплив повноти нечіткої бази знань на тривалість та точність навчання.

### Література

1. Штовба С.Д. Проектирование нечетких систем средствами MATLAB / Штовба С.Д. – М.: Горячая линия – Телеком, 2007. – 288 с.
2. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / Леоненков А.В – СПб: БХВ-Петербург, 2005. – 736 с.
3. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилинский М., Рутковский Л. – М.: Горячая линия – Телеком, 2006. – 452 с.
4. Прикладные нечеткие системы/ Т. Тэрано, К. Асаи, М. Сугэно и др. – М. Мир, 1993. – 368 с.
5. Штовба С.Д. Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным / Штовба С.Д. // Проблемы управления и информатики. – 2007. – №4. – С. 102–114.
6. Ротштейн О.П. Проектування нечітких баз знань: лабораторний практикум та курсове проектування: навч. посіб. / Ротштейн О.П., Штовба С.Д. – Вінниця: Вінницький державний технічний університет, 1999. – 65 с.