УДК 658.012

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

ЗАЛЕЖНІСТЬ ТОЧНОСТІ ІДЕНТИФІКАЦІЇ ВІД ОБСЯГУ НЕЧІТКОЇ СИНГЛТОННОЇ БАЗИ ЗНАНЬ

Анотація: Досліджується задача ідентифікації багатофакторних залежностей за допомогою нечітких синглтонних баз знань. Запропонована квадратична модель оцінки точності бази знань від її повноти.

Аннотация: Исследуется задача идентификации многофакторных зависимостей с помощью нечетких синглтонных баз знаний. Предложенная квадратичная модель оценки точности базы знаний от ее полноты.

Abstract: The work is devoted to the problem of identifying multi-dependency by singleton fuzzy knowledge bases. The quadratic model is proposed for evaluating the accuracy of the knowledge base of its fullness.

Вступ

Ідентифікацію багатофакторних залежностей в техніці, економіці, медицині, соціології, будівництві, сільскому господарстві, спорті та в інших областях все частіше здійнюють на онові нечітких баз знань. Нечіткою базою знань називається сукупність нечітких правил “Якщо – тоді”, яка задає взаємозв'язок між входами та виходами досліджуваного об'єкту. Існує декілька моделей баз знань, такі як: Мамдані, синглтона, Сугено. В синглтонній базі знань антецеденти представлені нечіткими множинами, а консеквенти – дійсними числами. Алгоритми логічного виведення за синглтонною базою найпростіші, тому її достатньо часто застосовують в нечіткій ідентифікації.

Під час побудови математичної моделі за результатами спостережень вирішують задачі структурної та параметричної ідентифікації [1]. Задачі параметричної ідентифікації на основі синглтонної бази знань легко формалізуються у формі задач неперервної оптимізації [2 – 4]. Відповідні задачі структурної ідентифікації мало досліджені – на сьогодні залишаються відкритими питання про вибір кількості нечітких термів, виду функцій належності, об’єму бази знань тощо [3–8]. **Метою дослідження** є визначення впливу кількості правил на точність нечіткої синглтонної бази знань. Виявлення такої залежності дасть змогу спроектувати адекватну базу знань з мінімальною кількістю правил. що забезпечить її швидше навчання.

Постановка задачі

Вважатимемо відомою тестову вибірку з M пар експериментальних даних, що повязують фактори впливу $X=(x_1,x_2,...,x_n)$ з виходом y досліджуваної залежності:

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

де X_r – вхідний вектор в r-ому рядку вибірки та y_r – відповідний вихід.

Позначимо через $y=F(N,\, X)$ – модель на основі нечіткої синглтонної бази знань з N нечітких правил, що пов’язують X з y. Точність ідентифікації визначимо як середньоквадратичну нев’язку на вибірці (1):

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

За фіксованого нечіткого розбиття вхідних та вихідної змінних можна згенерувати кілька нечітких баз знань 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 для трьох еталонних залежностей з подальшою апроксимацією аналітичними моделями.

Нечітке виведення за синглтонною базою знань

Синглтонну нечітку базу знань запишемо так [3]:

$$\left(x_1=\tilde{a}_{1j}\ \ ma\ \ x_2=\tilde{a}_{2j}\ \ ma\ldots ma\ \ x_n=\tilde{a}_{nj}\right)\Rightarrow y=d_j,\quad j=\overline{1,\, N}\ ,$$

де $\tilde{a}_{ij}$ – нечіткий терм, яким оцінено фактор x_i в j-му правилі, $i=\overline{1,n}$, $j=\overline{1,N}$;

d_j – консеквент j-го правила, який задано дійсним числом.

Ступіні належності поточного вхідного вектора $X^*=\left(x_1^*,\, x_2^*,\ldots,\, x_n^*\right)$ до числових значень $d_1, d_2, ..., d_m$ розраховують так:

$$\mu_{d_j}(X^*) = \mu_j(x_1^*) \cdot \mu_j(x_2^*) \cdot \ldots \cdot \mu_j(x_n^*)) \;,\quad j = \overline{1, N}\,,$$

де $\mu_j(x_i^*)$ – ступінь належності значення x_i^* нечіткому терму $\tilde{a}_{ij}$;

Чітке значення на виході моделі розраховують через дефаззіфікацію нечіткої множини $\tilde{y} = \left(\frac{\mu_{d_1}(X^*)}{d_1}, \frac{\mu_{d_2}(X^*)}{d_2}, \ldots, \frac{\mu_{d_m}(X^*)}{d_m} \right)$ за методом центра тяжіння:

$$y = \frac{\sum_{i=1}^{m} d_i \mu_{d_i}(X^*)}{\sum_{i=1}^{m} \mu_{d_i}(X^*)}\,.$$

Комп’ютерні експерименти

Експерименти проведемо для 3 еталонних залежностей (рис. 1) – неспадної, унімодальної та багатоекстремальної:

$$y = x_1\sqrt{x_2}\,,\quad x_1 \in [2; 22],\ x_2 \in [2; 14], \tag{3}$$

$$y = -x_1^{\ 2} - x_2^{\ 2}\,,\ x_1 \in [-7; 3],\ x_2 \in [-5; 5], \tag{4}$$

$$y = (1 + \sin(x_1)^2)^{x_2}\,,\ x_1 \in [0; 5],\ x_2 [0.5; 2]\,. \tag{5}$$

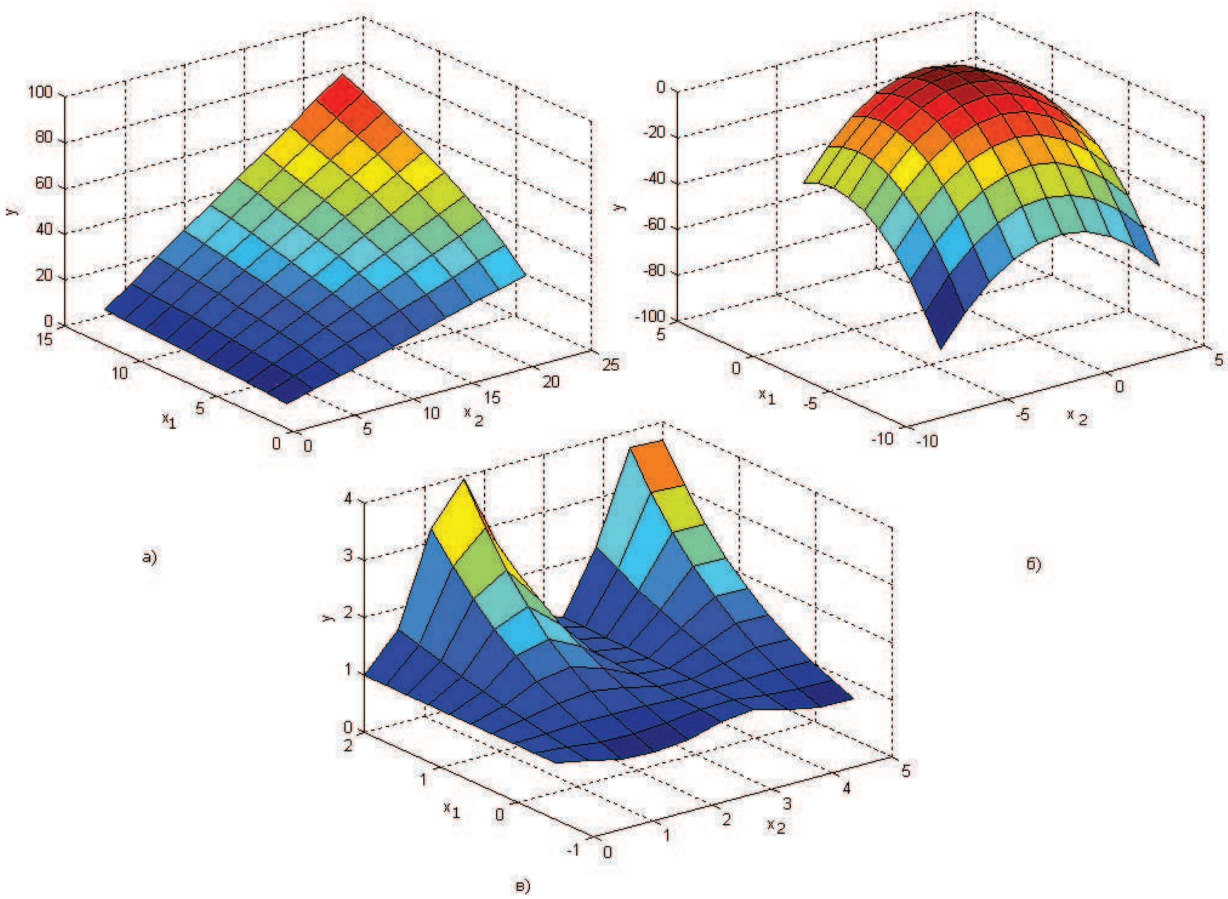

Рисунок 1 – Еталонні залежності: а) неспадна (3); б) унімодальна (4); в) багатоекстемальна (5)

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 $RMSE_{mean}$ нев’язки;
6) побудувати графіки залежностей $RMSE_{\min}$, $RMSE_{mean}$ та $RMSE_{\max}$ від N.

Нечітке розбиття здійснено за допомогою гаусових функцій належності, ядра яких рівномірно розподілено на діапазоні вхідних змінних. Коефіцієнт концентрації функцій належності прийнято рівним с=Δ***core*** /2.4, де Δ***core*** – відстань між ядрами сусідніх термів. За такого розподілу висота перетину нечітких множин сусідніх термів дорівнює 0.5. Консеквент кожного правила розраховувався за формулами (3) – (5) для ядер нечітких термів антецедента.

Для неспадної залежності (3) для оцінки вхідних змінних використовувалось 2, 3 та 4 термів, тобто експерименти проведено для таких 9-ти нечітких розбиттів вхідних змінних: 2х2, 2х3, 2х4, 3х2, 3х3, 3х4, 4х2, 4х3 та 4х4. Відповідно, максимальна кількість адекватних нечітких правил (N_{max}) склала 4, 6, 8, 6, 9, 12, 8, 12 та 16. Таким чином, 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. Для унімодальної залежності (4) використовувалися такі нечіткі розбиття вхідних змінних: 3х3, 3х4, 3х5, 4х3, 4х4 та 5х3. Для багатоекстремальної залежності (5) використовувалися такі нечіткі розбиття вхідних змінних: 4х2, 4х3, 4х4, 5х2 та 5х3.

Результати експериментів (рис. 2) показали, що нев’язка RMSE_{mean} спадає зі збільшенням кількості нечітких правил і досягає мінімума за повної бази знань. Якщо вдало підібрати комбінацію правил, тоді нев’язка RMSE_{min} стає суттєво меншою за RMSE_{mean}. В багатьох випадках мінімум RMSE_{min} 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 [8]. Бази знань з цього «плато насичення» назвемо прийнятними. Планку для них призначимо у вигляді 50% перевищення нев’язки у порівнянні з найкращою нечіткою моделі. На рис. 3 наведені пари таких моделей за нечіткого розбиття 4х4:

– для неспадної еталонної залежності (3) найкраща база знань містить 16 правил, а прийнятна – 6;
– для унімодальної еталонної залежності (4) найкраща база знань містить 12 правил, а прийнятна – 7;
– для багатоекстемальної еталонної залежності (5) найкраща база знань містить 16 правил, а прийнятна – 8.

Зведемо результати усіх експериментів до одного масштабу. Для цього для кожної бази знань розрахуємо відносне відхилення її нев’язки від нев’язки найкращої бази знань. Позначимо цю величину через *delta*_RMSE. Результати виконання цих дій наведено на рис. 4. Проведена квадратична апроксимації експериментальних даних (рис. 4) свідчить, що повну нечітку базу знань можна скоротити в 2–3 рази без великих втрат точності. Бази знань, які містять біля 70-80% від максимальної кількості правил, мають найвищу точність.

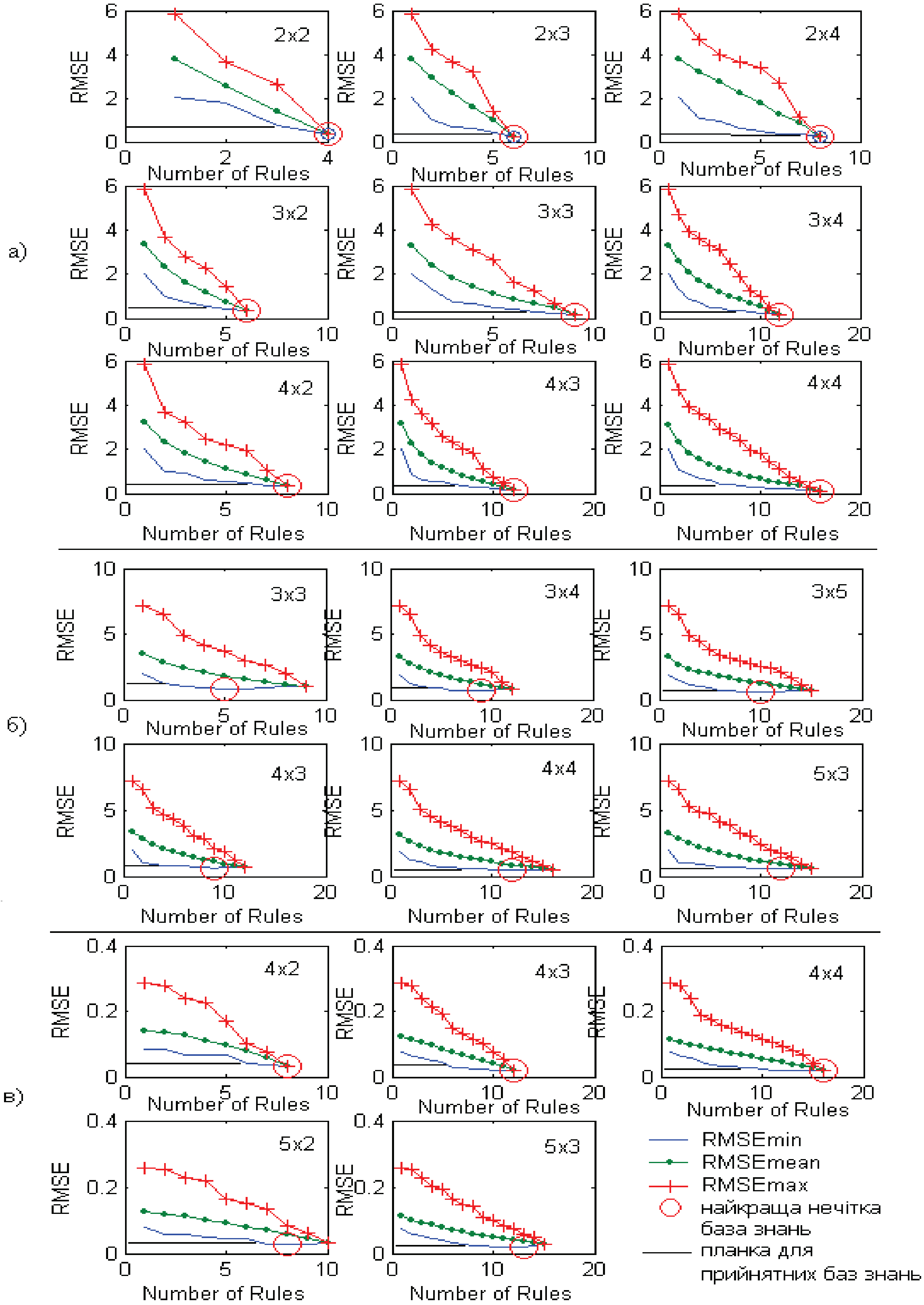

Рисунок 2 – Експериментальні криві навчання нечітких синглтонних баз знань:

а) для залежності (3); б) для залежності (4); в) для залежності (5)

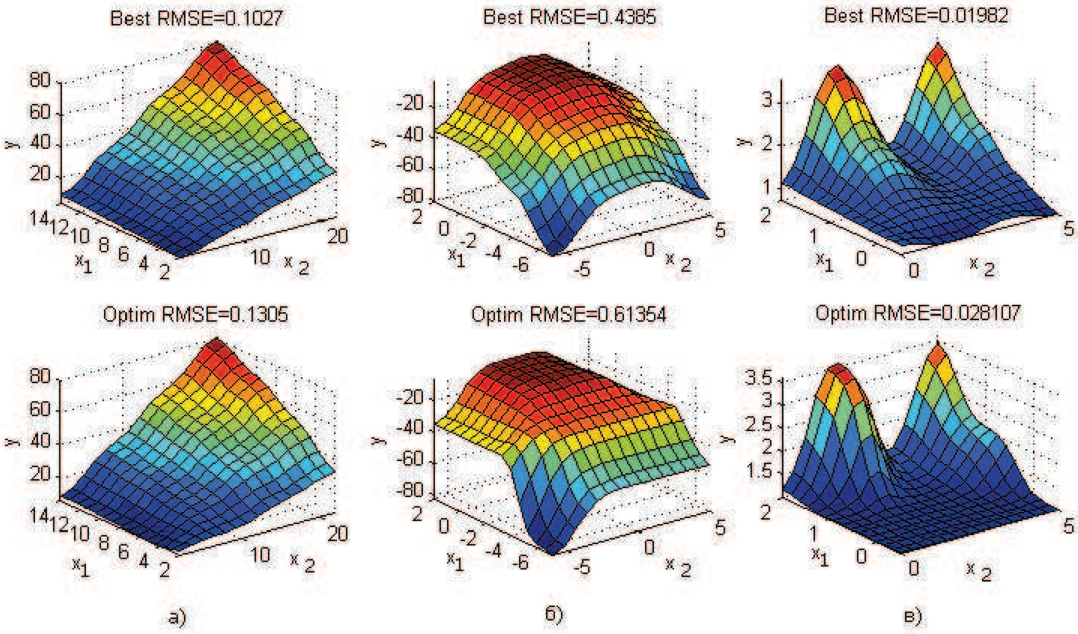

Рисунок 3 – Поверхні найкращої та прийнятної синглтонної бази знань за нечіткого розбиття 4×4 а)для залежності (3); б)для залежності (4); в)для залежності (5)

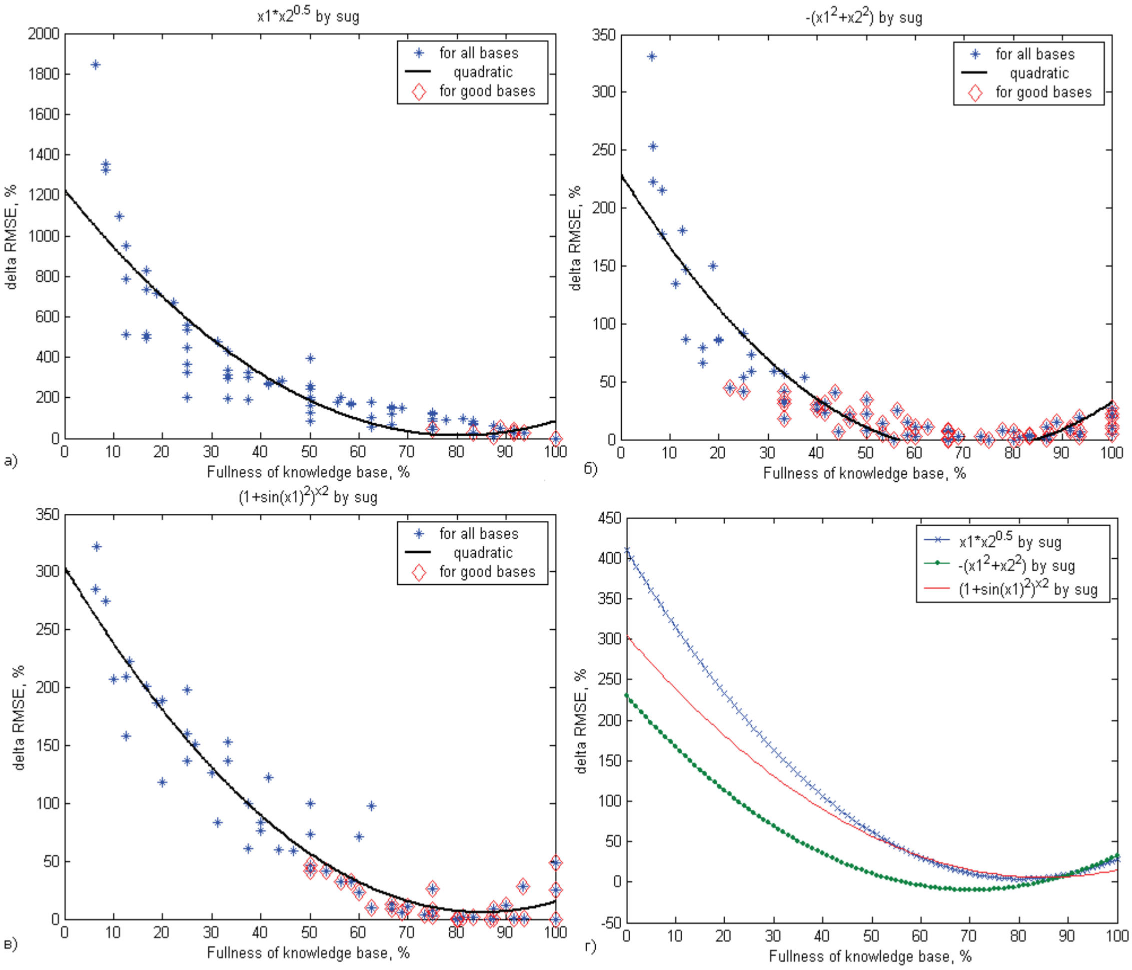

Рисунок 4 – Залежність приросту похибки від повноти нечіткої синглтонної бази знань: а) для залежності (3); б) для залежності (4); в) для залежності (5); г) для всіх залежностей

Висновки

Проведені комп'ютерні експерименти показують, що в більшості випадків найменшу похибку ідентифікації дає неповна нечітка синглтонна база знань, що наповнена правилами на 70-80%. На нашу думку, це обумовлено деякими суперечливостями на границях нечіткого розбиття через взаємодію великої кількості правил. Встановлено, що залежність точності нечіткої бази знань від її розмірності може бути описана поліномом другого порядку. 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 для інших форматів нечітких баз знань. Крім того, варто дослідити вплив повноти нечіткої бази знань на тривалість та точність навчання.

Література

1. Цыпкин Я.З. Основы информационной теории идентификации / Цыпкин Я.З. – М.: Наука, 1984. – 320 с.
2. Takagi T. Fuzzy Identification of Systems and Its Applications to Modeling and Control / Takagi T., Sugeno M. // IEEE Trans. on Systems, Man, and Cybernetics. – 1985. Vol. 15, №1. – P. 116-132.
3. Штовба С.Д. Проектирование нечетких систем средствами MATLAB / Штовба С.Д. – М.: Горячая линия – Телеком, 2007. – 288 с.
4. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети / Ротштейн А.П. – Винница: УНІВЕРСУМ–Вінниця, 1999. – 320 с.
5. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилинский М., Рутковский Л. – М.: Горячая линия – Телеком, 2006. – 452 с.
6. Zimmermann H.-J. Fuzzy Sets Theory and Its Applications / Zimmermann H.-J. 3rd ed. – Kluwer Academic Publisher, 1996. – 435 p.
7. Сергиенко М.А. Методы проектирования нечеткой базы знаний / Сергиенко М.А. // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2008. – № 2. – С. 67-71.
8. 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.

Відомості про авторів

Штовба Сергій Дмитрович – доцент, д.т.н., професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет.

Панкевич Ольга Дмитрівна – доцент, к.т.н, доцент кафедри теплогазопостачання, ВНТУ

Мазуренко Віктор Володимирович – студент, факультет автоматики та комп'ютерних систем управління ВНТУ.