

(Intercept) Rank YOY_Units
 19.1062345 -0.3048700 0.8149075

Рисунок 1 – відібрані змінні та їх оцінки за методом відбору оптимальної підмножини

(Intercept) (Intercept) Rank Sales Units
 24.0998436999 0.0000000000 -0.3589370601 -0.0589067560 0.0560364111

YOY_Units Unit_Volume Franchising
 0.6990154040 -0.0001751964 -0.1891396162

Рисунок 2 - відібрані змінні та їх оцінки на основі гребневої регресії

(Intercept) (Intercept) Rank Sales Units YOY_Units
 18.6115042354 0.0000000000 -0.2499263234 0.0000000000 0.0009136236 0.7807299523

Unit_Volume Franchising
 0.0000000000 0.0000000000

Рисунок 3 - відібрані змінні та їх оцінки на основі метода ласо

Таблиця 1 – показники точності отриманих моделей

Назва моделі	Відбір оптимальної підмножини	Гребнева регресія	Метод ласо
MAE(Середня абсолютна похибка)	122,56	130,55	124,94

Як показано на рис.1 оптимальна модель містить дві змінні: Rank та YOY_Units, в даному прикладі наглядно продемонстровано недолік гребневої регресії та перевага метода ласо. А саме: в гребневу модель потрапили всі змінні, деякі з них з малими коефіцієнтами, що значно ускладнює її інтерпретацію, а от в модель ласо потрапило три змінні, це зменшує складність моделі, та покращує її інтерпретацію, також, з табл.1 видно, що вона має кращий показник точності.

Список літератури:

1. Gareth James, Trevor Hastie, Robert Tibshirani, Daniela Witten – «An Introduction to Statistical Learning: with Applications in R» – Springer, 2013. 456p.
2. В.В. Христиановский, Т.В. Нескорородева, Ю.Н. Поликов Экономико-математические методы и модели: практика применения в курсовых и дипломных работах. Донецк, 2012.
3. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Електронний ресурс]. – Режим доступу до ресурсу: <https://ranalytics.github.io/data-mining/042-Regularization.html> (дата звернення 15.04.2022)
4. Хмелівський Ю.С Нескорородева Т.В. Аналіз даних для прогнозування серцевої недостатності засобами мови R. Матеріали II всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Комп'ютерні технології обробки даних" (10 грудня 2021 року) - Вінниця: ДонНУ імені Василя Стуса., с.57-60.
5. Restaurant Business Rankings 2020 [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.kaggle.com/datasets/michau96/restaurant-business-rankings-2020> (дата звернення 13.04.2022)

УДК 004.82

Резнік Р.Ю., студент 1 курсу спеціальності «Комп'ютерні технології обробки даних (DataScience)»
 Штовба С.Д., д-р. техн. наук, професор, професор кафедри інформаційних технологій

СИНТЕЗ НЕЧІТКОЇ БАЗИ ЗНАНЬ МАМДАНІ З ВИКОРИСТАННЯМ СУПЕРЕЧЛИВИХ ПРАВИЛ

Донецький національний університет імені Василя Стуса, м. Вінниця

Нечіткою базою знань називається сукупність нечітких правил “Якщо – тоді”, якими описується залежність між входами та виходами досліджуваного об’єкту. Часто для нечіткого моделювання використовують базу знань Мамдані, в якій антецеденти та консеквенти задаються нечіткими множинами [1]. Під час синтезу нечіткої бази знань Мамдані вважаються, що вони несуперечливі, тобто відсутні правилами з однаковими антецедентами та різними консеквентами, наприклад, такі правила, як:

Якщо (x_1 = низький та x_2 = середній), тоді y = середній,

Якщо (x_1 = низький та x_2 = середній), тоді y = високий.

Подивимося на суперечливі правила під іншим кутом зору. Якщо під час логічного висновку виконується антецедент таких суперечливих правил, то для формування вихідної нечіткої множини подається і консеквент першого правила, і консеквент другого правила. В результаті агрегування, отримуємо нечітку множину, яка буде між двома консеквентами, в нашому прикладі буде між середнім та високим, тобто відповідатиме терму вище середнього. А такого терма немає в терм-множині вихідної змінної y може і не бути. Таким чином, застосування суперечливих правил може підвищити так звану роздільну здатність опису залежності, яку моделює база знань. Відповідно, з’являються шанси підвищити точність нечіткої бази знань за рахунок використання суперечливих правил. Для перевірки цієї гіпотези проведемо обчислювальні експерименти з ідентифікації залежностей за допомогою нечітких баз Мамдані з суперечливими правилами.

Експерименти проведемо за методикою з роботи [2] для таких двофакторних залежностей (рис. 1):

$$y = x_1^2 - x_2^3 * \tan(0.1 * x_1) \in [-20, 25], x_2 \in [-1, 1.3]; \quad (1)$$

$$y = \frac{x_1^2 + x_2}{0.1e^{x_1} + x_2}, x_1 \in [-5, 5], x_2 \in [3, 7]; \quad (2)$$

$$y = x_1^2 - x_2^3 * \tan(0.1 * x_1) \in [-20, 25], x_2 \in [-1, 1.3]. \quad (3)$$

Для кожної з залежностей синтезовано множину нечітких правил Мамдані.

Для залежності (1) застосуємо лінгвістичне розбиття вхідних 4×4 , для залежності (2) – 4×3 , для залежності (3) – 4×5 . Відповідно, для цих залежностей синтезуємо 16, 12 та 20 несуперечливих правил-кандидатів. Далі, побудуємо криві навчання (рис. 2) у формі залежності RMSE від кількості відібраних до бази знань правил. RMSE розраховано на тестовій вибірці із 100 точок. Криві навчання побудовано для найкращого набору правил, який визначався жадібним алгоритмом.

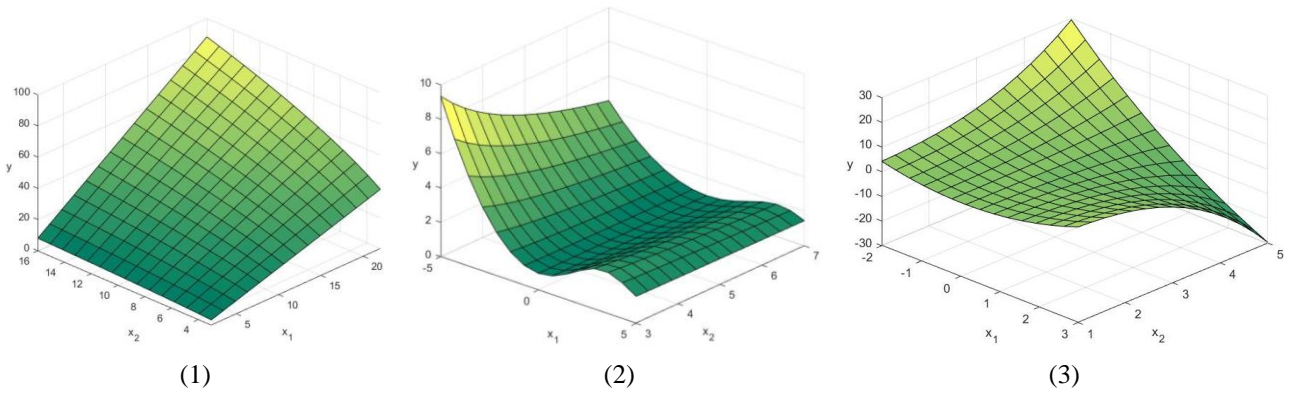


Рисунок 4 – Поверхні еталонних залежностей

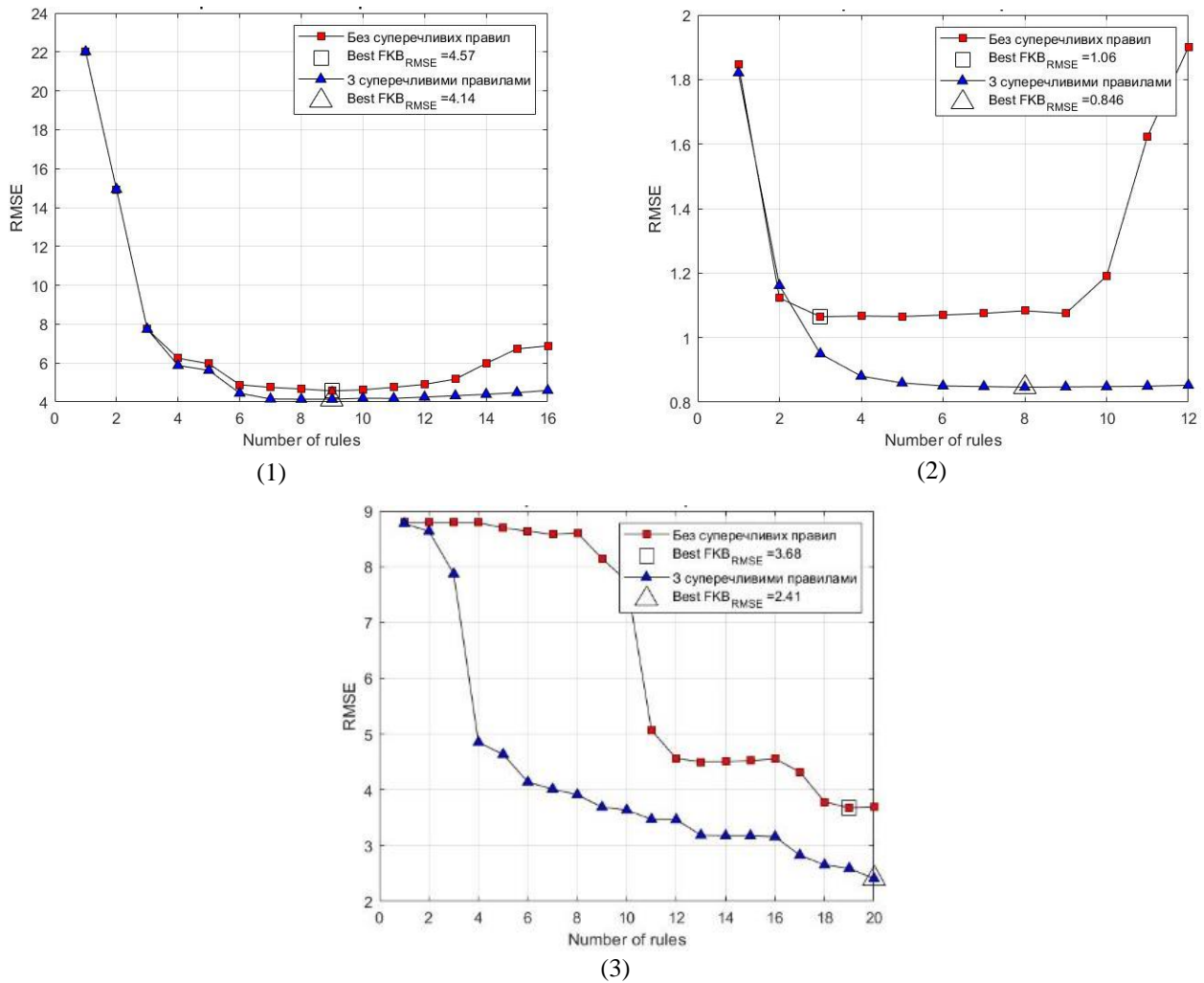


Рисунок 5 – Криві навчання баз знань з та без суперечливих правил

Ідентифікацію з використанням множини суперечливих правил здійснено у такий спосіб. Спочатку, збільшимо множину правил-кандидатів, додавши до кожного правила, ще одне конкурентне. Консеквентом конкурентного правила призначимо терм, що має другий ранг у алгоритмі генерування правила Ванга-Менделя [3]. Порівнюючи криві навчання на рис.2, на основі суперечливих та несуперечливих множин правил-кандидатів, бачимо доцільність використання суперечливих правил для нечіткої бази знань Мамдані. В усіх експериментах

нечіткі бази знань з суперечливими правилами кращі за точністю ніж звичайні нечіткі бази знань. Для близької до лінійної залежності (1) покращення несуттєве. Для суттєво нелінійних залежностей (2) та (3) покращення точності за рахунок використання суперечливих правил значне.

Для порівняння на рис. 3 для залежності (3) виведено поверхні нечітких баз знань одного обсягу з несуперечливими та з суперечливими правилами. З цього рисунка видно, що нечітка база знань з суперечливими правилами не лише має менше $RMSE$, але і візуально краще відновила залежність.

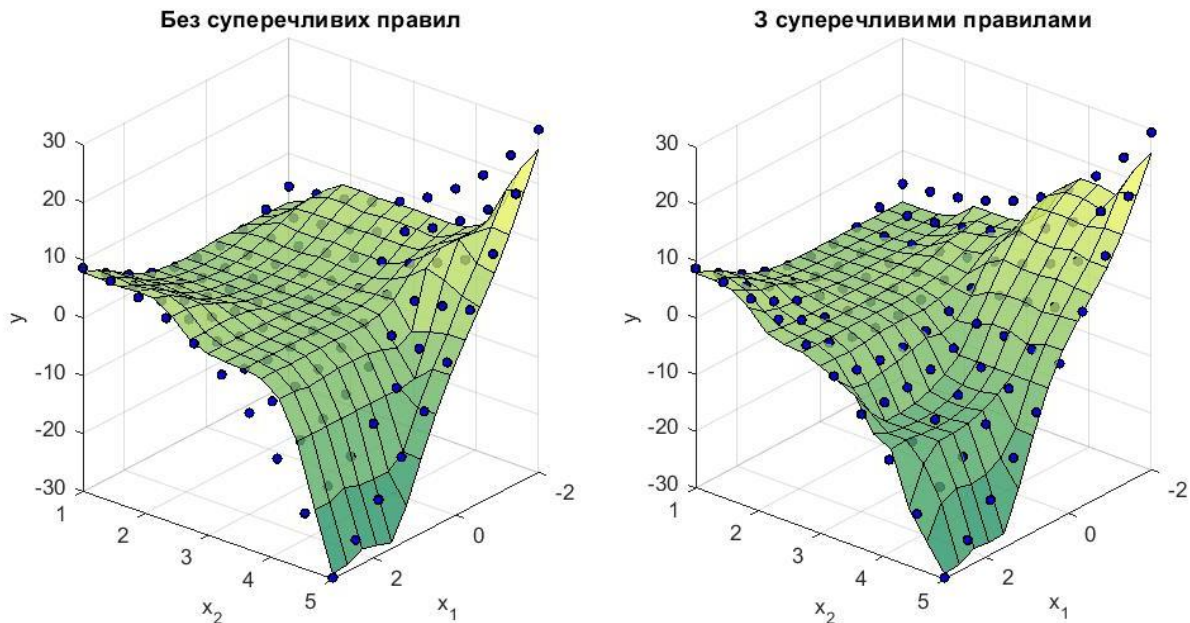


Рисунок 6 - Поверхні найкращих баз знань (3) залежності, з нанесеними на них тестовими значеннями.

Запропонований підхід, використання суперечливих правил в базі знань Мамдані, забезпечує суттєво кращу точність, за рахунок підвищення роздільної здатності опису. Це дозволяє досягнути бажаної точності нечіткої бази знань за меншої кількості правил.

Список літератури:

1. Штовба С.Д., Галушак А.В. Ідентифікація багатофакторних залежностей за допомогою баз знань. Лабораторний практикум: електронний навчальний посібник. – Вінниця: Вінницький національний технічний університет, 2016. – 96 с.
2. Штовба С.Д. Вплив кількості нечітких правил на точність бази знань Мамдані/ Штовба С.Д., Мазуренко В.В., Панкевич О.Д. // Вісник Хмельницького національного університету. Технічні науки. – 2011. – №2. – С. 185–188.
3. Wang L. X., Mendel J. M. Generating fuzzy rules by learning from examples //IEEE Transactions on systems, man, and cybernetics. – 1992. – T. 22. – №. 6. – С. 1414-1427.