

НЕЧІТКА СИСТЕМА АНАЛІЗУ ДАНИХ ДЛЯ КРЕДИТНОГО СКОРИНГУ ФІЗИЧНИХ ОСІБ

Виконав: ст. гр. 1КН-16спм Лавренюк А.О.

Науковий керівник: к.т.н., проф. Месюра В. І.

ОБ'ЄКТ, ПРЕДМЕТ І МЕТА ДИПЛОМНОЇ РОБОТИ

Об'єктом дипломної роботи є процес ідентифікації нечітких моделей кредитного скорингу фізичних осіб.

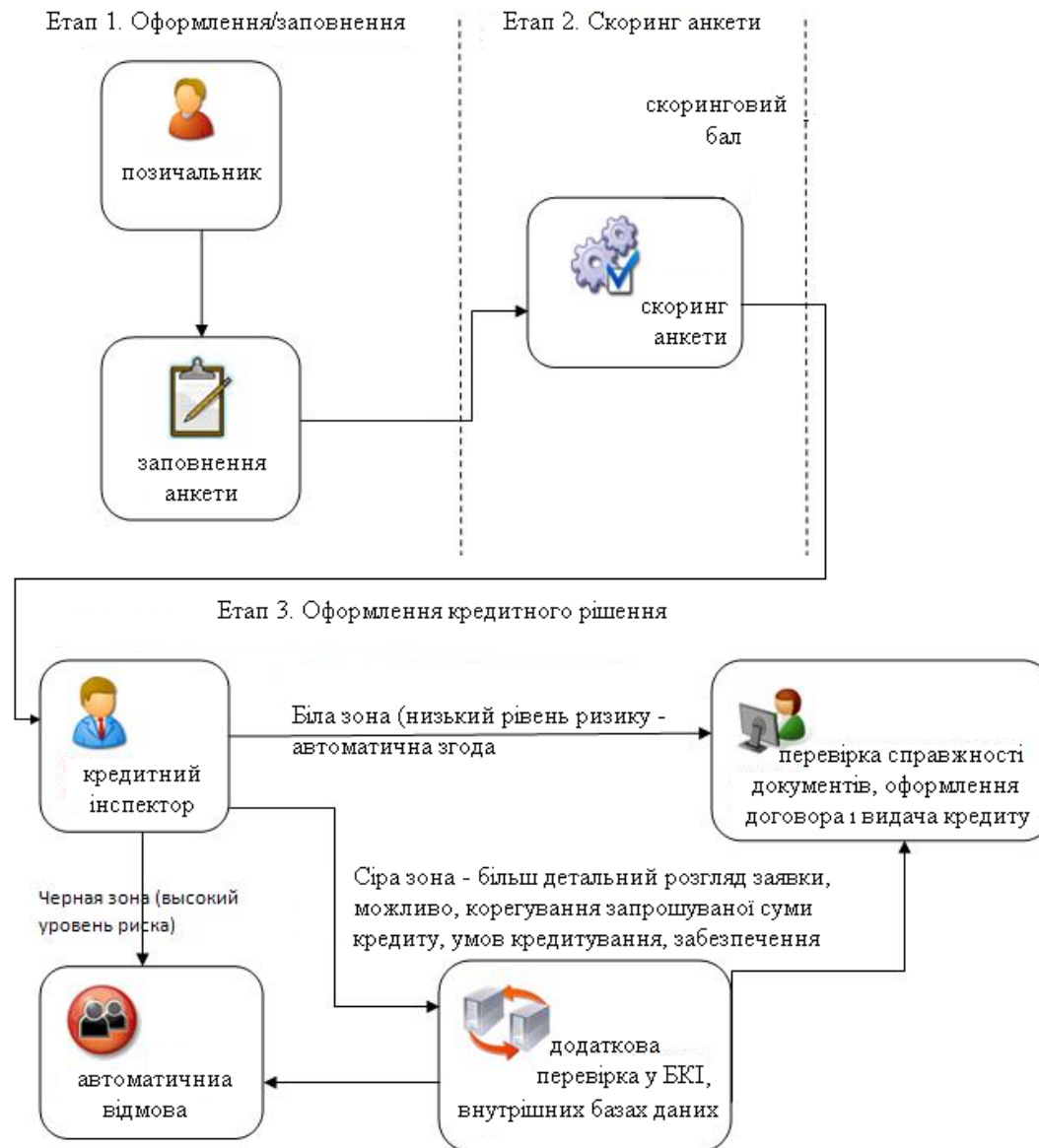
Предметом роботи є комплекс алгоритмів і програм ідентифікації антецедентів і консеквентів нечітких правил.

Метою роботи є підвищення точності та прозорості нечіткої системи аналізу даних для кредитного скорингу.

Основними задачами роботи є:

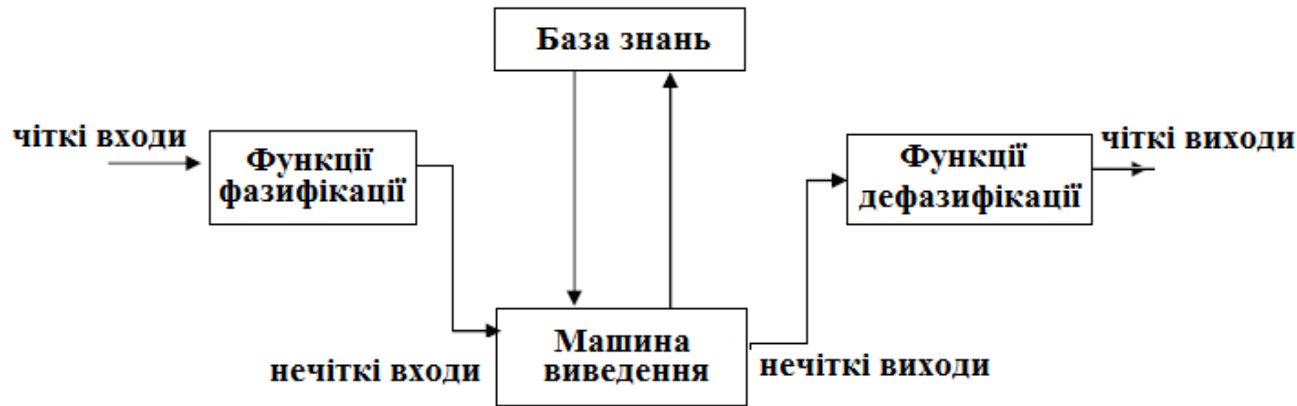
- аналіз предметної області для вибору найбільш ефективних методів розв'язання задачі аналізу даних кредитного скорингу з метою визначення основних елементів його агрегованої моделі;
- аналіз методів машинного навчання, нечіткого інтелектуального аналізу даних та нечіткої логіки для визначення моделей, які будуть агрегуватися для підвищення ефективності кредитного скорингу;
- визначення порядку з'єднання окремих моделей, унікальні особливості яких забезпечать виявлення необхідних закономірностей у масивах даних кредитних історій, у єдиний агрегат;
- побудова механізму прийняття рішень агрегатом моделей на основі рішень окремих моделей, що входять до його складу;
- розробки алгоритмів та програмної реалізації процедури кредитного скорингу;
- здійснення тестування і порівняння отриманих результатів на «стандартних» наборах кредитних історій, які використовуються для дослідження точності методів, використовуваних у кредитному скорингу.

ТИПОВІ БІЗНЕС-ПРОЦЕСИ СКОРИНГУ ФІЗИЧНИХ ОСІБ



ІДЕНТИФІКАЦІЯ НЕЧІТКИХ МОДЕЛЕЙ

Нечітке логічне виведення



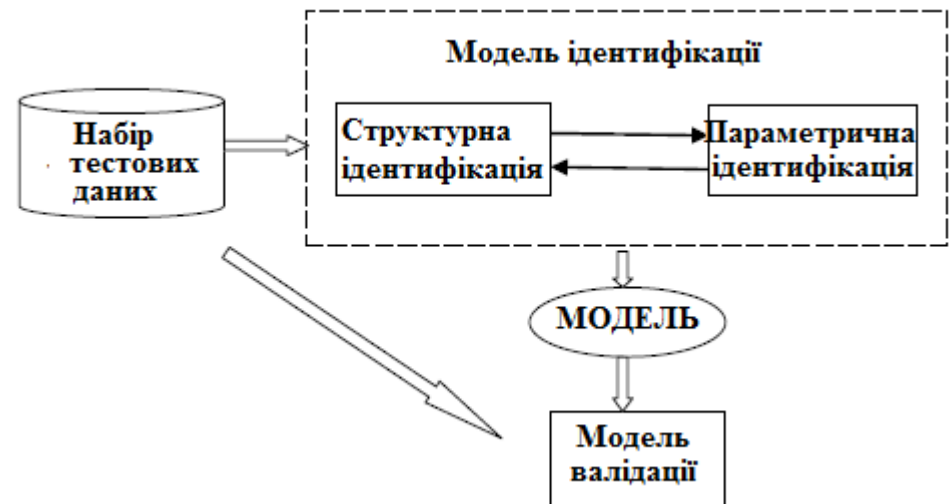
Задача моделювання скорингової систем

Цілі прогностичного моделювання:

- точність прогнозування;
- інтерпритованість результатів;
- адаптивність.

Основні етапи нечіткого логічного виведення:

- фазифікація чітких змінних (перетворення чітких змінних на нечіткі множини);
- отримання значення функції належності;
- нечітке логічне виведення (імплікація);
- агрегування (створення одного нечіткого набору для кожної вихідної змінної);
- дефазифікація (перетворення нечіткої множини на чіткий висновок).



НЕЧІТКІ МОДЕЛІ СУГЕНО І МАМДАНІ

Нечітка модель Сугено

$$R_k: \text{ЯКЩО } (x = A^k) \text{ТО } (y = h_j^k(x))$$

R_k – k -е правило бази знань; x – вхідна змінна;

A^k – лінгвістичний терм, що подає нечітку множину, яка описує x ;

$y = h_j^k(x)$ – вихідна змінна, $h_j^k(x)$, $j = 1, \dots, m$ – поліноміальні функції входів, подають локальні моделі апроксимації відгуку системи в області вхідного простору, поданого антецедентом A_k .

$$\mu_{ik}(x_i) = \exp\left(-\frac{(x_i - c_{ik})^2}{2a_{ik}^2}\right)$$

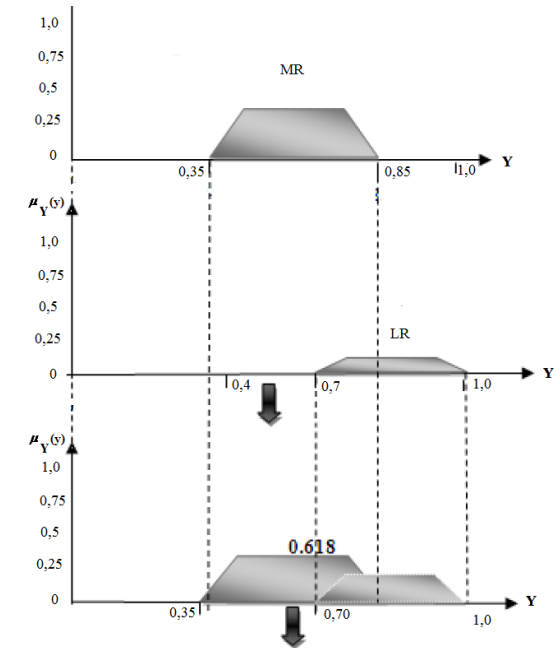
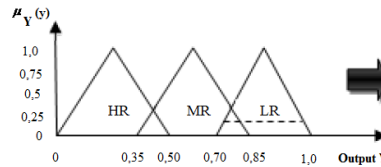
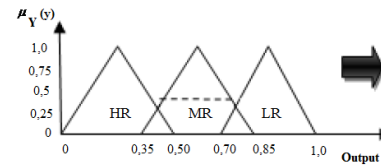
$\mu_{ik}(x_i)$ – функція належності;

c_{ik} і a_{ik} – параметри центру і ширини відповідних лінгвістичних термів.

Нечітка модель Мамдані

$$R_k: \text{ЯКЩО } (x = A^k) \text{ТО } (y = B_j^k)$$

де R_k – мітка k -го нечіткого правила, A^k і B_j^k – відповідно, нечіткі вхідні і вихідні множини.



ПРОЗОРІСТЬ. ПОКАЗНИКИ ПРОЗОРОСТІ



Зменшення кількості правил у базі знань



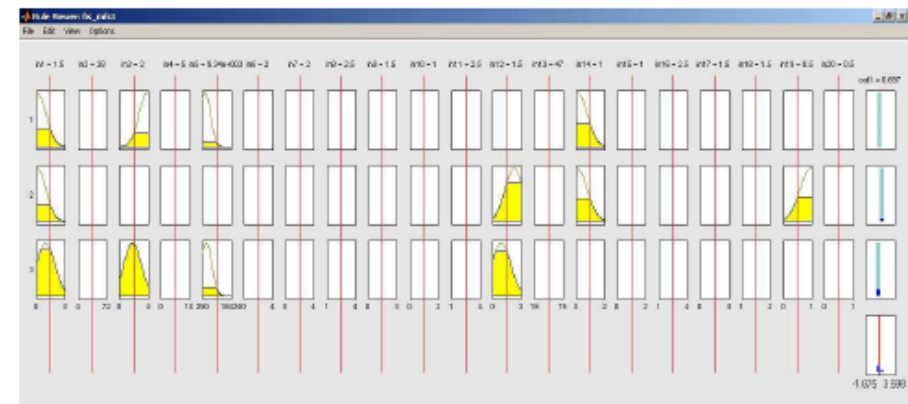
Приклад продукційного правли

(ЯКЩО Сума кредиту є Дуже низька) І (ЯКЩО Ощадний рахунок є Невідомо) і (ЯКЩОІ інші поручителі є Гарант) і (ЯКЩО Вік є Дуже низький) і (ЯКЩО Інші розстрочки є Банк) і (ЯКЩО Житло є безкоштовно) (ТО Клієнт ПОГАДННИЙ)

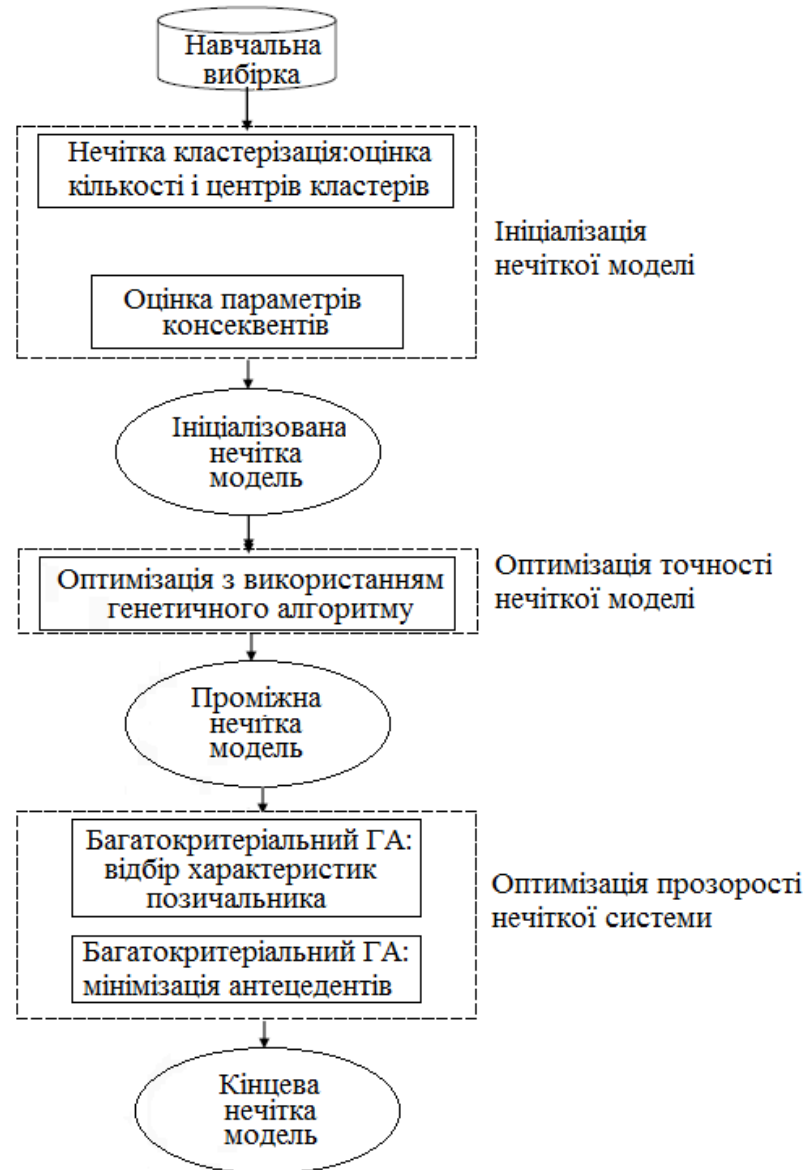
Зменшення кількості вхіих параметрів



Зменшення кількості значень лігністичних змвних



ЗАБЕЗПЕЧЕННЯ ПРОЗОРОСТІ НЕЧІТКОЇ СИСТЕМИ СУГЕНО



ВИКОРИСТАННЯ НЕЧІТКИХ ГЕНЕТИЧНИХ АЛГОРИТМІВ ДЛЯ ІДЕНТИФІКАЦІЇ НЕЧІТКИХ СИСТЕМ СУГЕНО

Особина популяції:

$$S_i = (c_{11}^{(t)}, a_{11}^{(t)}, \dots, c_{n1}^{(t)}, a_{n1}^{(t)}, \dots, c_{1k}^{(t)}, a_{1k}^{(t)}, \dots, c_{nk}^{(t)}, a_{nk}^{(t)})$$

Визначення випадкових параметрів термів особин початкової популяції

$$c_{ik} = rand \times range\{x_i(t)\} + \min_{t=1..n}\{x_i(t)\}$$

$$a_{ik} = rand \times \frac{2}{3} \frac{range\{x_i(t)\}}{K_S} + a_{i,min} \quad range\{x_i\} = [\max_{t=1..n}\{x_i(t)\} - \min_{t=1..n}\{x_i(t)\}]$$

Значення фітнес-функції:

$$p_R(s_i) = \frac{fit(s_i)}{\sum_{j=1}^N fit(s_j)}$$

Фітнес-функція для відбору характеристик потенційного позичальникв:

$$Fitns(S_i) = fitn_{acc}(S_i) + a \times fitn_{input}(S_i)$$

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
1	1	0	0	1	0	0	1	1	1	0	0	1	1

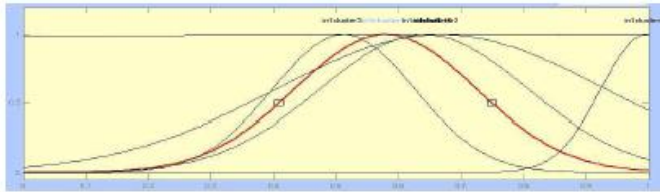
Фітнес-функція для відбору нечітких множин антецедентів:

$$Fitns(S_i) = fitn_{acc}(S_i) + a \times fitn_{mfs}(S_i)$$

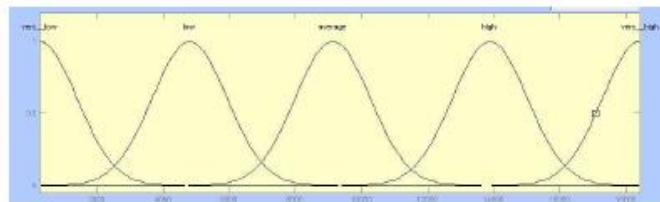
Правило 1									Правило K				
A_1^1	A_2^1	...	A_n^1	C_1	A_1^K	A_2^K	...	A_n^K	C_K
1	1	0	...	1	1	0	0	...	1

ІДЕНТИФІКАЦІЯ НЕЧІТКОЇ СИСТЕМИ МАМДАНІ

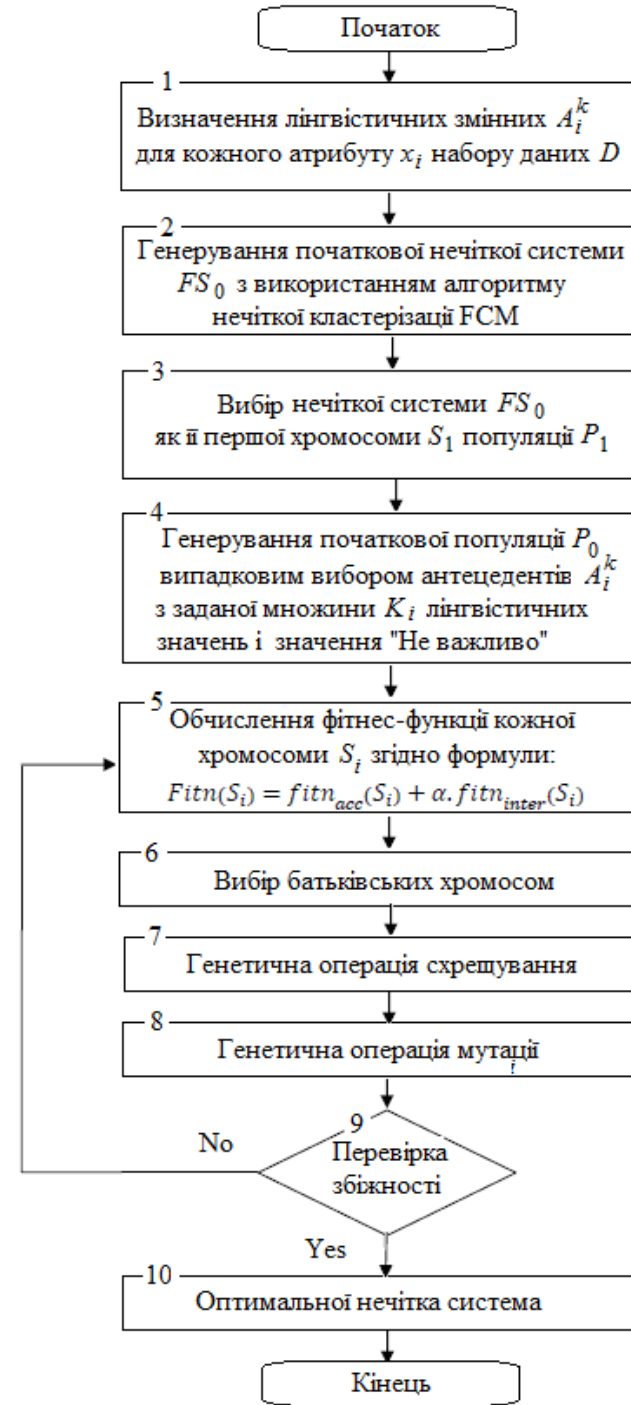
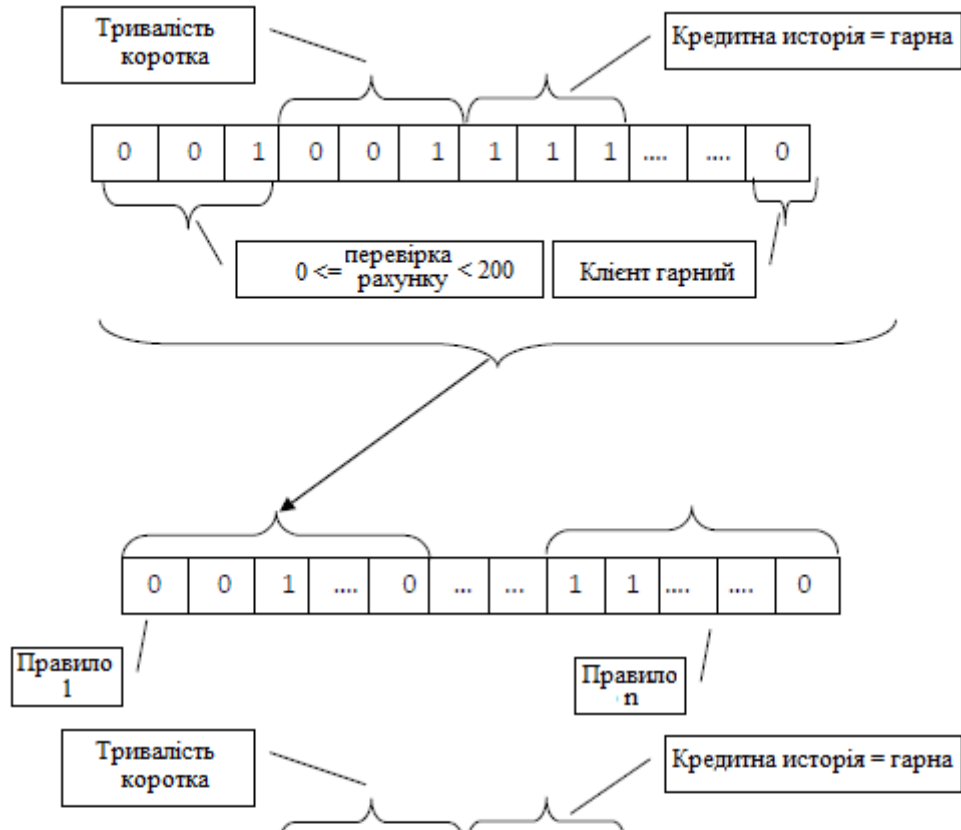
Результат автоматичної генерації правил



Регулярні інтерпретовані антецеденти



Застосування ГА до нечітких значень антецедентів



ТЕСТУВАННЯ ГІБРИДНИХ АЛГОРИТМІВ НЕЧІТКОЇ СИСТЕМИ АНАЛІЗУ ДАНИХ ДЛЯ КРЕДИТНОГО СКОРИНГУ

Нечітка система Мамдані		
Набір даних	Середнє (австр)	Середнє (нім)
Точність після створення початкової нечіткої системи	52,15	55,73
Точність після оптимізації ГА з чітко визначеними лінгвістичними нечіткими множинами антецедентів	87.91	77.88
Кількість антецедентів до використання БКГА	98	160
Кількість антецедентів після використання ГА	27	53.8
Кількість антецедентів «не важливо» після використання ГА	71	106.2
Кількість вилучених входів внаслідок використання ГА	2.2	2.4
Кількість антецедентів на одне правило після використання ГА	3.8	6.71

Нечітка система Сугено		
Набір даних	Середнє (австр)	Середнє (нім)
Кількість входних параметрів на одне правило до відбору входів	14	20
Нечітка кластеризація	87.34	72.80
Нечітка система з ГА оптимізацією	88.89	77.07
Кількість входних параметрів на одне правило до відбору входів	5.8	8.8
Відбір входів з БКГА	88.70	75.2
Кількість нечітких множин на одне правило до відбору входів	3	4.9
Нечіткі множини відібрані з використанням БКГА	88.60	75

АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ РОЗРОБЛЕНОЇ СИСТЕМИ

Метод	Австралійський набір	Німецький набір
Генетичний алгоритм + Машина опорних векторів (GA+SVM)	86.9%	77.92%
Генетичне програмування (GP)	88.27%	77.34%
Дерево класифікації та регресії (CART)	85.81%	70.59%
Алгоритм дерева рішень (C4.5)	87.06%	73.17%
Сирі множини (Rough sets)	83.72%	74.57%
Нейронні мережі (NNs)	88.3%	73.7%я
Радіально базисні функції	87.78%	75.63%
Алгоритми дипломної роботи:		
Нечітка система Сугено+ генетичний алгоритм (S Fuzzy + GA): оптимізація точності	88.89	77.07
Нечітка система Сугено + багатокритеріальний генетичний алгоритм (S Fuzzy+Multiobjective GA): оптимізація точності та прозорості	88.60	75
Нечітка система Мамдані + багатокритеріальний генетичний алгоритм (S Fuzzy+Multiobjective GA): оптимізація точності та прозорості	86.19	72.60

ПЕРСПЕКТИВИ ПОДАЛЬШОГО РОЗВИТКУ РОБОТИ

1. Розширення запропонованої моделі з моделі MISO (кілька входів - один вихід) на підтримку моделі MIMO (багато входів - багато виходів), щоб забезпечити можливість моделювання задач класифікації з кількома виходами;
2. Використання графічного інтерфейсу, подібного таблицям рішень для подання нечітких правил ЯКЩО - ТО, створених з наборів вхідних даних у додатковій, зручній для користувача, формі;
3. Випробування та перевірка запропонованих нечітких моделей на інших задачах класифікації, наприклад таких, як медична діагностика, і на більш загальних еталонних наборах даних, наприклад, таких як набори даних ірисів (Iris) і прогнозування раку молочної залози

ОСНОВНІ РЕЗУЛЬТАТИ ДИПЛОМНОЇ РОБОТИ

1. Сформульовані критерії вимірювання прозорості нечітких систем логічного виведення.
2. На основі багатокритеріального генетичного алгоритму розроблено методи відбору характеристик позичальника та зменшення кількості значень вхідних параметрів, що забезпечує збільшення показника прозорості при збереженні рівня точності прогнозування.
3. Запропоновані гібридні алгоритми автоматичного створення нечітких систем аналізу даних типу Сугено і Мамдіні, які на загальнодоступному наборі тестових даних довели свою перевагу за критеріями точності прогнозування і прозорості отриманих результатів.
4. Практичне значення одержаних результатів полягає у розробці програмного забезпечення аналізу даних для нечіткого кредитного скорингу на основі моделей Сугено і Мамдані, що дозволяє аналізувати дані з точки зору точності або прозорості отриманих результатів.
5. Розроблені алгоритми прийняті до впровадження у проекті скорингу нерухомості, що підтверджено довідкою про впровадження.
6. Попередні результати роботи апробовані на двох науково-технічних конференціях та опубліковані в електронних збірниках конференцій на сайті ВНТУ.

ДЯКУЮ ЗА УВАГУ!