УДК 519.85

ШОВГУН Н.В.

АНАЛІЗ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА ЗА ДОПОМОГОЮ МЕТОДІВ З НЕЧІТКОЮ ЛОГІКОЮ

В статті зроблено аналіз найбільш широко застосованих методів оцінки кредитоспроможності позичальника. Розглянуті методи оцінки кредитоспроможності на основі нечіткої логіки. Показані результати оцінки кредитоспроможності позичальника за допомогою методів з нечіткою логікою.

This paper is a survey on the most popular methods for assessing the creditworthiness of the borrower. Also, it presented methods for evaluating creditworthiness based on fuzzy logic and obtained results of assessing the creditworthiness of the borrower using methods of fuzzy logic.

1. Вступ

Основним напрямом діяльності комерційного банку є видача кредитів. По деяким оцінкам кредитування дає майже половину прибутку банку.

Разом з тим, кредитування пов'язано з ризиком, зумовленим можливим невиконання своїх обов'язків позичальником. Відомо, що цей ризик є один з найзначніших ризиків банку.

Споживче кредитування фізичних осіб є одним з основних банківських продуктів. При цьому для мінімізації втрат банку потрібен ретельний відбір позичальників та ефективна оцінка їх кредитоспроможності. Для цього в банках існують моделі і методи прийняття рішень по кредитним заявкам.

Насамперед, для прийняття рішення має значення детальна анкета позичальника, у якій містяться багато даних – від матеріального положення позичальника до його особистих якостей. З врахуванням усіх цих даних прийняте рішення повинно мінімізувати ризик та одночасно не повинно мати наслідком необґрунтовану втрату позичальника.

Умови, у яких має бути прийняте рішення, часто характеризується неповнотою даних, їх різнорідністю та недостатністю.

Метою даної статті є дослідження існуючих підходів до визначення кредитоспроможності позичальника, аналіз їх особливостей та вибір методик, які найбільш ефективно дозволяють прийняти рішення у конкретних умовах діяльності комерційного банку. Запропоновано використовувати для аналізу кредитоспроможності фізичних осіб нечіткий контролер Мамдані та нечітку нейрону мережу TSK з механізмом нечіткого логічного висновку Сугено.

2. Аналіз відомих методів оцінки кредитоспроможності позичальника

Найчастіше для оцінки кредито-спроможності позичальника використовують:
- статистичні методи;
- дерева рішень;
- генетичні алгоритми;
- нейронні мережі.

Статистичні методи на основі дискримінантного аналізу використовуються при вирішенні задач класифікації. Так, відома модель Альтмана, що використовується для аналізу кредитоспроможності юридичних осіб, побудована на основі множинного дискримінантного аналізу. Для аналізу кредитоспроможності фізичних осіб використовують лінійну або логістичну регресії [1]. Наприклад, при використанні лінійної регресії, функція, що визначає кредитний рейтинг, апроксимується лінійною функцією щодо компонентів вектора характеристик позичальника, тобто

$$p = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots a_N \cdot x_N ,$$

де a_0 – вільний член;

a_i при $i = 1, \ldots, N$ – вагові коефіцієнти характеристик позичальника;

x_i – характеристики позичальника.

Всі регресійні методи чутливі до кореляції між характеристиками, тому в моделі не повинно бути сильно корельованих характеристик позичальника.

Статистичні методи спираються на усереднені характеристики вибірки, але при дослідженні реальних складних життєвих феноменів ці характеристики можуть не відповідати дійсності. Оцінка кредитоспроможності позичальника за допомогою цих методів потребує великої кількості історичних данних щодо кредитів,

що не завжди можливо. Також характерною є проблема недостатньої кількості прикладів позичальників, що виявилися не спроможними погасити свою заборгованість. Одне з вирішень цієї проблеми запропоновано в [2]. Суттєвим недоліком статистичних методів вважають вимоги до спеціальної підготовки користувача [3].

Метод дерев рішень відрізняється високою швидкістю обробки даних і навчання при збереженні властивостей систем нечіткого логічного висновку.

При використанні методу дерев рішень для класифікації кредитних заявок застосовується набір правил, що формується при побудові дерева на основі навчальної вибірки. Дерево включає взаємопов'язані початковий (кореневий), проміжні та кінцеві вузли Кожному з вузлів відповідає умова (правило) класифікації об'єктів. Для побудови дерева на кожному внутрішньому вузлі необхідно знайти таку умову, яка б розбивало множину, асоційовану з цим вузлом на підмножину. В якості такої перевірки повинен бути вибрана одна з характеристик. Обрана характеристика повинна розбити множину так, щоб одержані в підсумку підмножини складалися з об'єктів, що належать до одного класу, або були максимально наближеними до цього, тобто кількість об'єктів з інших класів в кожній з цих множин було якомога менше.

Але у метода дерев рішень існують суттєві недоліки. Він не підходять для задач з великим числом можливих розв’язків і умова (правило) може формулюватися тільки в термінах «більше/менше», що заважає застосуванню до задач, де класс визначається більш складним поєднанням змінних [4].

Генетичні алгоритми грунтуються на стохастичному пошуку глобального оптимума цільової функції. Ідея генетичних алгоритмів запозичена в живої природи і полягає в організації еволюційного процесу (за допомогою операцій схрещування, мутації та селекції), кінцевою ціллю якого є оптимальне рішення задачі.

Генетичні алгоритми мають ряд недоліків. Критерій відбору хромосом і сама процедура відбору евристичні і не гарантують знаходження «кращого» рішення. Також необхідно мати в наявності досить великий обсяг вхідних данних для завершення процедури селекції [5].

Під нейронними мережами розуміють обчислювальні структури, які моделюють прості біологічні процеси, що зазвичай асоціюють з процессами в людському мозку. Вони являють собою совокупність елементів (штучних нейронів), пов'язаних між собою синаптичними зв'язками [6].

Недоліком нейронних мереж називають те, що вони являють собою «чорний ящик» та відсутність твердих правил, щодо вибору швидкості навчання мережі для вирішення конкретних задач [6].

Існує багато архітектур нейронних мереж. Так наприклад, в роботі [7] для класифікації клієнтів німецького та австралійського банків використовувалися такі мережі: мережа Кохонена, мережа BackPropagation, радіально-базисна мережа, каскадна мережа.

Загальними недоліками вищеописаних методів є вимоги до об’єму вхідних даних та жорсткі вимоги до характеристик і критеріїв відбору позичальника. В реальному житті середовище позичальника постійно змінюється, як результат – змінюються його сімейний та фінансовий стан, що вносить невизначеність в інформацію щодо клієнта. Неправильна оцінка позичальника в цих умовах може призвести до збільшення ризику банку чи втрати потенційно надійних клієнтів.

Врахувати таку невизначеність можна за допомогою методів нечіткої логіки. Даний підхід дає можливість працювати як з кількісними, так і з якісними характеристиками.

3. Методи оцінки кредитоспроможності позичальника за допомогою нечіткої логіки

Задачу оцінки кредитоспроможності можна сформулювати таким чином. Кожна кредитна заявка задається вектором $\{X_1, X_2, ..., X_i, ..., X_N\}$, де X_i – певним чином формалізовані дані з анкети позичальника та параметри кредиту. Далі по заданому вектору треба прийняти рішення про надання кредиту, тобто класифікувати позичальника як «надійного», чи класифікувати як «поганого».

В статті [8] використовувався нечіткий контролер з алгоритмом нечіткого логічного висновку Мамдані для аналізу кредитоспроможності юридичних осіб. В загальному вигляді нечіткий логічний висновок має наступні етапи:

1. Визначення множини вхідних змінних:
$X = \{X_1, X_2, ..., X_i, ..., X_M\}$;
2. Визначення множини вихідних змінних:
$D = \{D_1, D_2, ..., D_i, ..., D_M\}$;

3. Формування базової терм-множини з відповідними функціями належності кожного терма: $A = \{a_1, a_2, ..., a_i\}$;

4. Формування кінцевої множини нечітких правил, узгоджених щодо використовуваних в них змінних;

5. Знаходження чіткого значення для кожної з вихідних лінгвістичних змінних.

Даний підхід можна використати і для розв'язання задачі оцінки кредитоспроможності фізичних осіб.

Розглянемо алгоритм нечіткого висновку Мамдані.

Нехай базу знань складають два нечітких правила:

$П_1$: якщо x є A_1 і y є B_1, то z є C_1,

$П_2$: якщо x є A_2 і y є B_2, то z є C_2,

де x і y вхідні змінні, z – вихідна змінна, A_1, A_2, B_1, B_2, C_1, C_2 – деякі задані функції належності, при цьому чітке значення z треба визначити на основі даної інформації і чітких значень x, y. Етапами алгоритму є:

1. Введення нечіткості. Знаходимо степінь істиності для передумов кожного правила: $A_1(x_0)$, $A_2(x_0)$, $B_1(x_0)$, $B_2(x_0)$.

2. Логічне виведення. Знаходимо рівні «відсікання» для передумов кожного з правил (з використанням операції мінімума):

$$\alpha_1 = A_1(x_0) \cap B_1(y_0);$$

$$\alpha_2 = A_2(x_0) \cap B_2(y_0).$$

Далі знаходимо «відсікання» функції належності:

$$C_1' = (\alpha_1 \cap C_1(z));$$

$$C_2' = (\alpha_2 \cap C_2(z)).$$

3. Композиція. Знаходимо об'єднання знайдених відсічених функцій належності з використанням операції максимум, отримуємо підсумкову нечітку підмножину для змінної виходу з функцією належності:

$$\mu_\Sigma = C(z) = C_1'(z) \cup C_2'(z) = (\alpha_1 \cap C_1(z)) \cup (\alpha_2 \cap C_2(z)).$$

4. Зведення до чіткості з використанням, наприклад, центроідного методу [8].

В алгоритмі нечіткого висновку Сугено, використовується такий набір правил [9]:

$П_1$: якщо x є A_1 і y є B_1, то $z=a_1x+ b_1y$,

$П_2$: якщо x є A_2 і y є B_2, то $z=a_2x+ b_2y$,

де x і y вхідні змінні, z – вихідна змінна, A_1, A_2, B_1, B_2, C_1, C_2 – деякі задані функції належності, a_1, a_2, b_1, b_2 – деякі числа.

Алгоритм має вигляд:

1. Введення нечіткості як в алгоритмі Мамдані.

2. Нечітке виведення. Знаходимо $\alpha_1 = A_1(x_0) \cap B_1(y_0)$, $\alpha_2 = A_2(x_0) \cap B_2(y_0)$ та індивідуальні виходи правил:

$$\dot{z}_1 = a_1 x_0 + b_1 y_0;$$

$$\dot{z}_2 = a_2 x_0 + b_2 y_0.$$

3. Визначення чіткого значення змінної виведення:

$$z_0 = \frac{\alpha_1 \dot{z}_1 + \alpha_2 \dot{z}_2}{\alpha_1 + \alpha_1}$$

Розглянемо нечітку нейронну мережу TSK (Takagi, Sugeno, Kang'a) з механізмом нечіткого логічного висновку Сугено. Правила мережі можна представити у такому вигляді [9]:

$$R_1 : якщо \quad x_1 \in A_1^{(1)}; x_2 \in A_2^{(1)}, ..., x_n \in A_n^{(1)},$$

$$то \; y_1 = p_{10} + \sum_{j=1}^{N} p_{1j} x_j;$$

$$R_M : якщо \quad x_1 \in A_1^{(M)}; x_2 \in A_2^{(M)}, ..., x_n \in A_n^{(M)}, то \; y_M = p_{M0} + \sum_{j=1}^{N} p_{Mj} x_j,$$

де $A_i^{(k)}$ – значення лінгвістичної змінної x_i для правила R_k з функцією належності

$$\mu_A^{(k)}(x_i) = \frac{1}{1 + \left(\frac{x_i - c_i^{(k)}}{\sigma_i^{(k)}} \right)^{2b_i^{(k)}}}.$$

Композиція результатів має вигляд:

$$y(x) = \frac{\sum_{k=1}^{M} w_k y_k(x)}{\sum_{k=1}^{M} w_k},$$

де $w_k = \mu_A^{(k)}(x)$ – степінь виконання умов правила, $\mu_A^{(k)}(x) = \prod_{j=1}^{N} \left[\frac{1}{1 + \left(\frac{x_j - c_j^{(k)}}{\sigma_j^{(k)}} \right)^{2b_j^{(k)}}} \right].$

5. Результати досліджень

Для аналізу кредитоспроможності позичальника за допомогою ННМ TSK та контролера Мамдані використовувалися дані одного з вітчизняних банків по 100 кредитним заявкам за 2010 р.

У якості вхідних змінних використовувався вектор X= {Сума кредиту, Дохід, Термін кредиту, Вік, Термін проживання в квартирі, Освіта}.

Співвідношення навчальної та перевірочної вибірки було вибрано як 70 до 30. Результати досліджень зведені в таблиці 1-2. Результати оцінки кредитоспроможності позичальника за допомогою лінійної та логістичної регресії, кластерного аналізу методом нечітких к-середніх [9] та вище наведених методів представлені в таблиці 3.

Табл. 1. Результати оцінки кредитоспроможності позичальника за допомогою ННМ TSK

Кількість правил	СКО навчальної вибірки	% невірних класифікацій на навчальної вибірці	СКО перевірочної вибірки	% невірних класифікацій на перевірочній вибірці
3	0.0316	2.857	0.0178	6.66
5	0.0149	0	0.0127	3.33
7	0.0194	0	0.027	3.33

Табл. 2. Результати оцінки кредитоспроможності позичальника за допомогою контролера Мамдані

Кількість правил	СКО навчальної вибірки	% невірних класифікацій на навчальної вибірці	СКО перевірочної вибірки	% невірних класифікацій на перевірочній вибірці
3	0.08349	6.66	0.1092	10
5	0.0193	2.857	0.0305	6.66
7	0.0131	1.4286	0.0382	6.66

Табл. 3. Порівняльні результати оцінки кредитоспроможності позичальника різними методами

Метод	СКО навчальної вибірки	% невірних класифікацій на навчальної вибірці	СКО перевірочної вибірки	% невірних класифікацій на перевірочній вибірці
Лінійна регрессія	0.0987	8.57	0.04334	13.33
Кластерний аналіз		8.57		16.66
Логістична регерсія	0.0171	2.857	0.0344	10
ННМ TSK	0.0149	0	0.0127	3.33
НК Мамдані	0.0193	2.857	0.0305	6.66

Як можна бачити з аналізу таблиць, ННМ TSK та контролер Мамдані дають майже однаково задовільні рішення при кількості правил 3-7. Процент невірних класифікацій при застосуванні статистичних методів аналізу тієї самої вибірки надто високий.

6. Висновки

1.Основною проблемою в аналізі кредитоспроможності фізичної особи виступає неточність даних, недостатня база знань про минуле клієнтів та необхідність працювати з інгвістичними характеристиками, які важко піддаються математичній обробці.

2. В специфічних умовах функціонування вітчизняних банків методи з використанням нечіткої логіки, закрема методи на базі нечіткої нейронної мережі TSK та контролера Мамдані дають непогані результати, які дозволяють звести відсоток незадовільних рішень до 3…10%.

Перелік посилань

1. Руководство по кредитному скорингу / под ред. Элизабет Мэйз;– Минск: Гревцов Паблишер, 2008г. – 458с.
2. Паклин Н.Б., Уланов С.В., Царьков С.В. Построение классификаторов на несбалансированных выборках на примере кредитного скоринга // Искусственный интеллект. – 2010г. - №3 – с. 528-534

3. А.А. Ежов, С.А. Шумкий Нейрокомпьютинг и его применения в экономике и бизнесе. – М.: МИФИ, 1998.- 224 с.
4. Тегеран Т. Программируем колективный разум. – СПб: Символ-Плюс, 2008. – 368 с.
5. Рутковская Д., Пилинський М., Рутковский Л. Нейронніе сети, генетические алгоритмі и нечеткие системы.– М.: Горячая линия -Телеком, 2006. – 452 с.
6. В.В. Круглов, М.И. Дли, Р.Ю. Голунов Нечеткая логика и искусственные нейронные сети. – М.: Физматлит, 2000. — 224 с.
7. D. Michie, D.J. Spiegelhalter, C.C. Taylor Machine Learning, Neural and Statistical Classification. 1994
8. Зайченко Ю.П. Оценка кредитних банковских рисков с использованием нечеткой логики// Intelligent Information and Engineering Systems. – 2008. - №13 – с. 190-200
9. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. – К.: Издательский дом «Слово», 2008. – 334с.