

ПРОГНОЗУВАННЯ РИЗИКІВ КРЕДИТУВАННЯ ФІЗИЧНИХ ОСІБ ЗА МАТЕМАТИЧНИМИ МОДЕЛЯМИ

© Кожухівська О.А., 2013

Розглянуто сучасні підходи до математичного моделювання процесу роздрібного кредитування. Також встановлена необхідність побудови спеціалізованих систем підтримки прийняття рішень на основі принципів системного аналізу під час кредитування фізичних осіб. Перевагами таких систем є можливість застосування ефективної попередньої обробки даних, використання декількох альтернативних методів оцінювання стану клієнтів та множини критеріїв якості на кожному етапі аналізу даних.

Ключові слова: роздрібне кредитування, математичні моделі, оцінювання стану клієнтів, прийняття рішень.

In the Paper modern approaches to mathematical modeling of retail crediting process have been considered. The necessity to construct specialized decision support systems on the basis of system analysis principles in crediting potential clients was recognized. Advantages of such systems are possibilities to implement effective preliminary data processing, use several alternative methods to estimate clients' status and set of quality criteria on each stage of data analysis.

Key words: retail crediting, mathematical models, client state estimation, decision making.

Вступ

Одним із основних напрямів діяльності банківських установ є кредитування фізичних осіб. Належне управління такими фінансовими процесами надає можливість встановити взаємовигідні довгострокові відносини між позичальниками кредиту та фінансовими установами. Значні накопичення банківських капіталів на сучасному етапі, розширення пропозицій банків та підвищення рівня організації та уніфікації бізнес-процесів сприяють підвищенню динаміки зростання кількості роздрібних клієнтів та обсягів виданих кредитів. Також спостерігається спрощення вимог банків до потенційних клієнтів і скорочення часу, необхідного для прийняття рішення стосовно можливості кредитування конкретної особи. Очевидно, що наявність впливу таких факторів призводить до зростання втрат внаслідок реалізації відповідних фінансових ризиків. Якщо темпи приросту та дохідність кредитного портфеля достатньо високі, то вони повністю перебивають фінансові втрати внаслідок реалізації ризиків. Саме тому більшість фінансових організацій протягом тривалого періоду не робили належних інвестицій в розроблення сучасних ефективних методів контролю кредитування та впровадження сучасних інформаційних комп'ютерних технологій, спрямованих на підтримку прийняття рішень під час управління ризиками роздрібного кредитування. Однак ситуація з роздрібним кредитуванням поступово змінилась у напрямі погіршення, особливо з настанням чергової світової фінансової кризи, незважаючи на її (в основному) зовнішнє походження.

Починаючи з осені 2008 року, більшість фінансових організацій зіткнулись з такими проблемами: 1) з'явилися істотні обмеження стосовно доступу до валютних фінансових ресурсів; відсутність ресурсів означає обмеження обсягів кредитування, а отже, і зменшення обсягів та темпів зростання дохідності кредитного портфеля; 2) внаслідок кризи у економічній сфері (зростання курсів валют, зниження рівня оплати праці, часткова або повна втрата роботи, зниження обсягів виробництва) істотно зменшились обсяги проблемних кредитів; 3) збільшилась кількість випадків шахрайства під час кредитування.

Все це свідчить про необхідність змінювати підходи до організації та супроводження процесу роздрібногo кредитування. Сьогодні існує гостра необхідність у створенні ефективних принципів управління фінансовими ризиками та надійних (з погляду результатів обчислень) комп'ютерних інформаційних систем підтримки прийняття рішень. Особливо актуальною стала потреба у застосуванні сучасних методів статистичного та інтелектуального аналізу даних, математичного моделювання фінансово-економічних процесів з метою побудови адекватних математичних моделей для прогнозування можливості повернення кредиту.

Загалом цикл кредитування складається з таких етапів [1–4]: 1) оцінювання кредитоспроможності клієнта; 2) супровід та моніторинг процесу сплати взятого кредиту; 3) здійснення заходів стосовно стягнення протермінованої заборгованості; 4) аналіз поточного стану кредитного портфеля і вироблення належних керуючих впливів; 5) постійне оновлення (адаптація) моделей оцінювання кредитоспроможності клієнтів до нових умов.

Тому в цій роботі розглянемо сучасні моделі й методи оцінювання кредитоспроможності фізичних осіб, на основі яких можна будувати комп'ютерні системи підтримки прийняття рішень з метою прискорення процесів аналізу даних і підвищення об'єктивності та якості рішень.

1. Аналіз існуючих розробок і постановка задачі

Мета роботи така: виконати аналіз сучасних методів побудови моделей для оцінювання кредитоспроможності фізичних осіб; вибрати методи моделювання для оцінювання кредитоспроможності клієнтів під час роздрібногo кредитування на основі характеристик позичальників; виконати обчислювальні експерименти стосовно прогнозування кредитоспроможності клієнтів і порівняти отримані результати.

Сучасні керуючі банківські інформаційні системи вже містять функції для обробки кредитних заявок (application processing system – APS). Це програмні комплекси для підтримки прийняття рішень стосовно кредитування. Фактично APS – це сучасний засіб для аналітичної підтримки реалізації кредитних процесів з використанням множини правил прийняття кредитних рішень. Такі програмні комплекси являють собою певний конструктор, впровадження якого означає отримання можливості повноцінної організації процесу видання кредитів. Однак цей комплекс має певні недоліки:

- функціональні обмеження в процесі видання кредитів: система містить лише ті параметри і механізми, які розробник вважає достатніми для ефективного управління ризиками; якщо замовнику необхідні нові функції та механізми, реалізувати їх можна лише шляхом модифікації та доповнення програмного коду.

- відносна складність впровадження нових процесів та реінжинірингу процесів; процес впровадження може тривати від чотирьох місяців до одного року.

На ринку існують програмні комплекси, у яких на певному рівні розв'язано згадані задачі, наприклад, система Experian. Подібні комплекси мають обмеження стосовно створення ієрархічних (багаторівневих) стратегій оцінювання кредитоспроможності позичальників. Наприклад, за допомогою такої системи можна розробити та впровадити правила дій залежно від кредитної історії, проте неможливо розробити сам механізм ідентифікації кредитної історії на основі наявної інформації стосовно платежів. Тобто такі комплекси не можуть виступати у вигляді єдиного модуля прийняття рішення, за допомогою якого аналітик зможе розробити або реалізувати процес реінжинірингу складних стратегій прийняття рішення.

Тому під час розв'язання практичних задач виникає необхідність проектування власних систем підтримки прийняття рішень на основі множини математичних моделей, інтегроване використання яких забезпечить ефективну високоякісну підтримку при прийнятті рішень стосовно кредитування фізичних осіб.

2. Математичні моделі для кредитного скорингу

Кредитний скоринг являє собою аналіз кредитоспроможності клієнтів. Процес прийняття рішень стосовно можливості видання кредиту позичальнику ґрунтується на знаннях і даних стосовно

клієнтів. Сучасні бази знань і даних (БЗД) інформаційних систем підтримки прийняття рішень містять низку означень, які сформулювали експерти з кредитування, спрямованих на пояснення значущості інформації, наявних моделей та можливостей їх використання у процесі кредитування. Ця інформація описує не тільки стандартні (відомі) можливості стосовно аналізу даних та вимоги до процесу кредитування, але й спеціальні знання (методи, моделі та обчислювальні алгоритми), їх інтерпретацію, внутрішню термінологію фінансової установи, яка стосується розв'язання задач, пов'язаних з кредитуванням.

Загалом моделі кредитного скорингу можна розділити на дві великі категорії: параметричні й непараметричні. Група параметричних моделей містить такі: 1) лінійні ймовірнісні моделі; 2) моделі типу логіт і пробіт (моделі двійкового вибору); 3) моделі, які ґрунтуються на застосуванні дискримінантного аналізу; 4) нейронні мережі; 5) нейронечіткі моделі; 6) байєсівські мережі. До непараметричних скорингових моделей належать: 1) моделі, які використовують для розв'язання задач математичного програмування; 2) класифікаційні дерева (рекурсивні класифікаційні алгоритми); 3) моделі, які використовують для реалізації методу найближчого сусіда; 4) аналітичний ієрархічний процес прийняття рішень; 5) нечітка логіка (і нечітка логіка в комбінації з іншими процедурами прийняття рішень); 6) експертне оцінювання і системи на його основі.

Лінійні ймовірнісні моделі. Лінійна ймовірнісна модель (ЛІМ) – це модель у формі лінійної регресії, залежна змінна якої набуває значення 0 або 1 залежно від того, яке рішення приймається стосовно видання кредиту. Формально така модель має вигляд:

$$y(k) = b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e(k), \quad (1)$$

де y – залежна змінна, значення якої відповідає результату прийняття рішення; $x_i, i=1, \dots, m$ – пояснювальні змінні (характеристики клієнта); b_i – коефіцієнти (параметри) рівняння регресії, які оцінюються за даними, які характеризують клієнтів; $e(k)$ – випадковий процес, зумовлений наявністю невимірюваних збурень, а також помилок (похибок) оцінювання структури і параметрів моделі; k – ідентифікатор клієнта. У векторній формі рівняння (1) має вигляд:

$$y = \mathbf{b}^T \mathbf{x} + e. \quad (2)$$

Отже, умовну ймовірність отримання кредиту можна записати так:

$$\text{Pr}(y | x) = \mathbf{b}^T \mathbf{x}. \quad (3)$$

Цю умовну ймовірність можна інтерпретувати як ймовірність отримання кредиту за умови інформації \mathbf{x} . Отже, після обчислення параметрів рівняння (1) його можна використовувати для оцінювання ймовірності надання кредиту новому клієнту. Отриману таким методом оцінку далі можна порівнювати з прийнятим граничним (пороговим) значенням з метою прийняття остаточного рішення стосовно видання кредиту. Застосування ЛІМ пов'язане з такими недоліками: 1) можливість набування залежною змінною значень, які перебувають поза інтервалом $[0, 1]$; 2) одночасне використання у правій частині категорійних змінних і змінних, представлених дійсними числами, може призводити до зміщення оцінок параметрів моделі; 3) процеси кредитування частіше характеризуються нелінійними залежностями, що потребує застосування моделей інших структур. Очевидно, що вказані недоліки можуть призводити до отримання неприйнятно грубих оцінок кредитоспроможності клієнта.

Нелінійні класифікаційні моделі логіт і пробіт. Для розв'язання задачі класифікації претендентів на отримання кредиту логічно використати функцію розподілу ймовірностей (кумулятивну функцію розподілу (КФР)), перетворену відповідно. КФР належать до класу монотонних функцій, тобто функцій, які монотонно зростають або спадають на деякому інтервалі. Припустимо, що для визначення ймовірності отримання кредиту p_c вибрано нормальний розподіл:

$$p_c = \Phi(\mathbf{b}^T \mathbf{x}) = \int_{-\infty}^u j(z) dz,$$

де $j(z)$ – щільність стандартного нормального розподілу; $u = \mathbf{b}^T \mathbf{x}$ – верхня границя інтегрування. Отже, можна отримати так звану модель пробіт.

Якщо для визначення ймовірності отримання кредиту вибрати функцію логістичного розподілу, то можна побудувати нелінійну модель логіт. Тобто у цьому випадку маємо:

$$p_c = \Phi(\mathbf{b}^T \mathbf{x}) = \int_{-\infty}^u j(z) dz = \frac{1}{1 + \exp(-\mathbf{b}^T \mathbf{x})}, \quad (4)$$

або

$$p_c = \frac{\exp(b_1 x_1 + \dots + b_m x_m)}{1 + \exp(b_1 x_1 + \dots + b_m x_m)}.$$

На відміну від функції нормального розподілу, логістична функція має замкнену форму, що забезпечує спрощення розрахунків за цією моделлю порівняно з моделлю пробіт. Параметри моделей обох типів оцінюють, як правило, за методом максимальної правдоподібності (ММП) без значних обчислювальних витрат. Альтернативним методом оцінювання є метод Монте-Карло для марковських ланцюгів (МКМЛ), який ґрунтується на генеруванні псевдовипадкових послідовностей (ПВП) і відборі випадкових значень, які задовольняють певні вимоги. Цей метод широко використовують для оцінювання нелінійних моделей завдяки наявності альтернативних методів генерування ПВП. Наприклад, його використовують у процедурах аналізу фінансових процесів для оцінювання моделей стохастичної волатильності [5]. Результати класифікації клієнтів, отримані за допомогою обох моделей, здебільшого прийнятні за якістю [1].

Дискримінантний аналіз. Постановка задачі у цьому випадку також вимагає розділення клієнтів на дві групи: G_1 – група клієнтів, які задовольняють поставлені вимоги, і G_2 – група, якій не надається кредит. Задача полягає у класифікації потенційних клієнтів на основі множини їх характеристик $\mathbf{x} = [x_1, x_2, \dots, x_m]$. Методика дискримінантного аналізу розв'язує цю задачу, обчислюючи дискримінантну функцію $I^T \mathbf{x}$, де I – функція вагових коефіцієнтів для компонент вектора \mathbf{x} . Значення вагових коефіцієнтів обчислюють, визначаючи максимально можливу різницю між обома групами клієнтів.

При цьому припускається, що вектор \mathbf{x} має нормальний розподіл для обох груп. Кожній групі можна поставити у відповідність пару значень (параметрів) (m_1, Σ_1) і (m_2, Σ_2) , які представляють групові середні й коваріації відповідно. Також треба ввести ймовірність p_i – ймовірність належності потенційного позичальника кредиту до групи i , і величину c_{ij} , яка характеризує втрати, пов'язані з неправильною класифікацією (коли претендента на кредит необхідно перемістити з групи i в групу j). У випадку, коли коваріаційні матриці обох груп збігаються, тобто $\Sigma_1 = \Sigma_2 = \Sigma$, тоді правило класифікації можна отримати з умови мінімізації вартості очікуваної неправильної класифікації. Ця постановка приводить до такого результату: претендент на кредит, який характеризується множиною значень \mathbf{x} , буде зарахований до групи G_1 , якщо

$$I^T \mathbf{x} \geq a + \ln \left(\frac{c_{21} p_2}{c_{12} p_1} \right), \quad (5)$$

де $I = \Sigma^{-1} (m_1 - m_2)$; $a = I^T (m_1 + m_2) / 2$.

В інших випадках претендента зараховують до групи G_2 .

Правило класифікації у цьому випадку дуже просте: значення, отримане за допомогою дискримінантної функції $I^T \mathbf{x}$, порівнюється з пороговим, яке визначається за виразом (5):

$$"norig" = a + \ln \left(\frac{c_{21} p_2}{c_{12} p_1} \right).$$

У разі перевищення порогу претендента зараховують до першої групи і навпаки. Вираз (5) називають моделлю лінійного дискримінантного аналізу, оскільки вектор \mathbf{x} входить в нього лінійно. У випадку, якщо $\Sigma_1 \neq \Sigma_2$, правило класифікації матиме квадратичну форму стосовно \mathbf{x} ; тому така модель належить до квадратичного дискримінантного аналізу.

Алгоритм (дерево) рекурсивної класифікації (АРК). Алгоритм рекурсивної класифікації (розділення) ґрунтується на правилі сортування, яке використовує послідовність двійкових підмножин описового критерію. Результатом застосування АРК є двійкове класифікаційне дерево, вузли та гілки якого утворюють структуру, що ставить у відповідність конкретне значення класифікаційної групи (G_1 або G_2) даним x_i потенційного клієнта. Наведемо просту ілюстрацію цього підходу.

Нехай необхідно розділити N суб'єктів на дві групи, G_1 і G_2 , використовуючи два критерії: **A** і **B**. Зарахування до однієї з груп виконується мінімізацією очікуваної вартості неправильної класифікації. Інакше кажучи, необхідно мінімізувати ризик необхідності змінити віднесення претендента до групи. Ризик присвоєння кінцевої точки t класифікаційного дерева групі G_1 можна формалізувати так:

$$R_1(t) = c_{21} p_2 p(2|t), \quad (6)$$

де $-p_i$ – ймовірність належності суб'єкта до групи i ; c_{ij} – вартість віднесення претендента на кредит до групи j , якщо він належить до групи i ; $p(2|t)$ – умовна ймовірність того, що суб'єкт, який належить до групи G_2 , буде присвоєний кінцевій точці дерева t . За аналогією, ризик присвоєння кінцевої точки t класифікаційного дерева групі G_1 описується так:

$$R_2(t) = c_{12} p_1 p(1|t).$$

Отже, якщо $R_1(t) < R_2(t)$, то алгоритм присвоїть кінцеву точку t класифікаційного дерева групі G_1 , а в інших випадках – групі G_2 .

Алгоритм рекурсивної класифікації розділяє вихідну множину даних (підвибірки) на дві частини на вершині класифікаційного дерева. Сортування даних виконується на основі концепції перемішаності вибірки за допомогою вибраного критерію або лінійної комбінації кількох критеріїв (характеристик). Як зазначено вище, $p(2|t)$ – умовна ймовірність того, що суб'єкт, який належить до групи G_2 , присвоюється кінцевій точці дерева t . Загалом можна визначити умовну ймовірність $p(i|t)$ стосовно групи i .

Формально ступінь перемішаності значення, присвоєного кінцевій точці t , можна записати так:

$$I(t) = R_1(t) p(1|t) + R_2(t) p(2|t).$$

Функцію $I(t)$ можна інтерпретувати як очікуваний ризик, який впливає з некоректної класифікації, тобто у кінцевій точці t суб'єкт належить до обох груп, а в дійсності його зараховують до групи i з ймовірністю $p(i|t)$. Інтегрований ступінь перемішування $I(T)$ для всього класифікаційного дерева T можна визначити як агреговану оцінку для всіх кінцевих точок.

Очевидно, що ступінь перемішування всієї вибірки у будь-якій точці дерева буде більшим, ніж ступінь перемішування окремих підвбірок, взятих з повної вибірки. Отже, логічно сформулювати таке правило класифікації у вузлі t , яке дасть можливість максимально зменшити ступінь перемішування даних. На основі цього АРК знаходить спочатку краще правило у заданій точці для кожної характеристики та їх комбінацій і на цій основі створює підвбірки. Така процедура бінарної класифікації продовжується доти, доки можна зменшувати ступінь перемішаності даних. На цьому процес класифікації закінчується і отримуємо класифікаційне дерево T_{\max} .

Останнім кроком АРК є вибір належного ступеня складності дерева виконанням його перехресної перевірки. Доволі часто дерева рішень, отримані за цим алгоритмом, мають високий ступінь складності, а тому ризик некоректної класифікації може бути значним. Тому необхідно виконувати перевірку побудованого дерева хоча б за допомогою частини даних, використаних у процесі побудови цієї моделі. Практичні застосування цього методу свідчать про його хороші класифікаційні характеристики, незважаючи на високу складність моделі. Це пояснюється його непараметричною природою.

Математичне програмування. Цей метод класифікації належить до непараметричних методів. Він надає більше можливостей для практичного застосування, ніж параметричні статистичні моделі, використання яких обмежено значеннями оцінок параметрів. Розглянемо задачу розділення даних на дві групи: G_1 і G_2 . Значення змінних і критерій класифікації для i -го суб'єкта містяться у векторі A_i . Задача полягає у визначенні такого вектора x і обмеження b , які задовольняють умову:

$$A_i x \leq b, \text{ якщо } A_i x \in G_1;$$

і

$$A_i x \geq b, \text{ якщо } A_i x \in G_2.$$

Дві групи розділяються гіперплощиною $Ax = b$. Якщо позначити через a_i ступінь порушення цієї умови суб'єктом, який характеризується вектором даних A_i , то для розв'язання задачі класифікації необхідно знайти

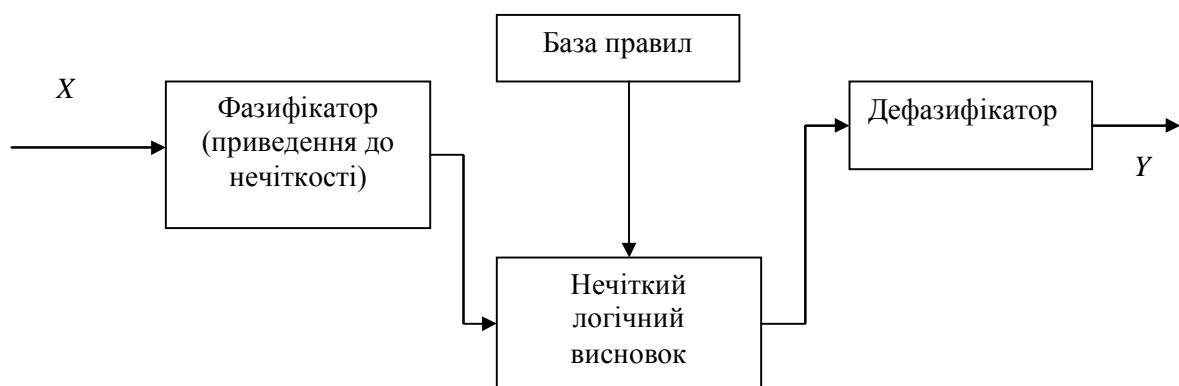
$$\min \sum_i c_i a_i, \quad (7)$$

де – $A_i x \leq b + a_i$, якщо $A_i x \in G_1$;

$A_i x \geq b - a_i$, якщо $A_i x \in G_2$.

Нечітка логіка. В умовах мінливих економічних відносин фінансові процеси функціонують під впливом випадкових збурень, що призводить до появи випадкових невизначеностей. Випадкова невизначеність означає, що деяка подія може відбутись або ні, тобто така подія є важко-прогнозованою. Але при цьому стан події вважається чітким і зрозумілим. Однак, разом із невизначеностями випадкового характеру, виникають невизначеності, зумовлені наявністю нечіткості. Невизначеність такого типу свідчить про нечіткість власне самого стану деякої події. Це призводить до того, що інтерпретація стану події може бути неоднозначною з боку різних дослідників, а це приводить до різних остаточних висновків. Тобто можна стверджувати, що невизначеність нечіткого типу може містити у собі суб'єктивну невизначеність. Загалом нечітку випадкову змінну можна розглядати як деяку вимірювану функцію, що набуває нечітких значень у ймовірнісному просторі нечітких змінних.

Теорія нечіткої логіки дає можливість отримати більшу кількість значень змінної, ніж експертний підхід, при цьому кожна змінна має визначену множину лінгвістичних значень. В загальному вигляді механізм формування логічного висновку складається з чотирьох етапів: введення нечіткості (фазифікація), формування нечіткого висновку, композиція і приведення до чіткості або дефазифікація (див. рисунок).



Процедура формування нечіткого логічного висновку

У випадку розв'язання задачі аналізу кредитоспроможності отримане за такою процедурою число характеризує кредитний рейтинг потенційного клієнта. Популярними алгоритмами формування нечіткого висновку є такі: Мамдані, Цукамото, Сугено та Ларсена, які відрізняються між собою виглядом правил, логічними операціями і методами дефазифікації.

Задачу оцінювання кредитоспроможності позичальника можна сформулювати так. Кожна кредитна заявка задається вектором $\{x_1, x_2, \dots, x_i, \dots, x_n\}$, де x_i – у певний спосіб формалізовані дані з анкети позичальника та параметри кредиту. Далі за заданим вектором вхідних даних потрібно прийняти рішення стосовно надання кредиту потенціальному клієнту, тобто оцінити вірогідність повернення кредиту. В загальному вигляді нечіткий логічний висновок складається з таких етапів: (1) – визначення множин вхідних $\{x_1, \dots, x_i, \dots, x_n\}$ і вихідних змінних $\{d_1, \dots, d_i, \dots, d_n\}$; (2) – формування базової терм-множини з відповідними функціями належності кожного терму: $A = \{a_1, \dots, a_i, \dots, a_n\}$; (3) – формулювання скінченної множини нечітких правил, узгоджених щодо змінних, які у них використовуються; (4) – знаходження чіткого значення для кожної вихідної лінгвістичної змінної. У цьому випадку за вихідну беремо лінгвістичну змінну «ймовірність повернення кредиту», яка має таку терм-множину: 1) дуже низька ймовірність повернення кредиту; 2) низька ймовірність повернення; 3) середня ймовірність повернення; 4) висока ймовірність повернення; 5) дуже висока ймовірність повернення кредиту.

Під час проведення аналітичних досліджень визначено перелік факторів, що впливають на формування рівня кредитоспроможності клієнта. Нижче наведено перелік деяких факторів, які використані для аналізу, зокрема такі: 1 – вік позичальника; 2 – стаж роботи; 3 – місячна заробітна плата; 4 – наявність майна; 5 – наявність кредитної історії; 6 – сума кредиту; 7 – термін кредитування. Отриманий результат наведено у табл. 1; за критерієм загальної точності нечітка логіка виявилась найкращою.

Байєсівські мережі (БМ). Байєсівська мережа – це модель ймовірнісного типу у формі спрямованого ациклічного графа, вершини якого – це вибрані змінні процесу, що моделюється. Кожній вершині ставиться у відповідність таблиця (апріорних) умовних ймовірностей, необхідна для розрахунків майбутніх станів вершин. Метою побудови такої моделі є встановлення причинно-наслідкових зв'язків між змінними, щоб уможливити формування ймовірнісного висновку, тобто умовної ймовірності настання подій, які нас цікавлять у конкретному випадку.

Для створення моделі у формі БМ необхідно розв'язати задачу структурного і параметричного навчання, тобто це класична задача математичного моделювання процесів довільної природи. Під час побудови БМ апіорну структуру мережі можна задати емпірично – тобто отримати її за експертними оцінками або за допомогою іншої інформації стосовно досліджуваного процесу. Якщо структура невідома, то її оцінюють, використовуючи наявну інформацію. За наявності статистичної інформації зручно використовувати евристичний алгоритм побудови БМ, який скорегує апіорну структуру мережі або дасть змогу побудувати таку структуру мережі, яку можна буде потім модифікувати, використовуючи знання експертів [6, 7]. Результатом застосування БМ є обчислення ймовірності неповернення кредиту за тих умов, яким відповідають значення інших змінних моделі. У табл. 1 наведено результати застосування трьох методів до аналізу кредитоспроможності позичальників кредиту (банківські дані).

Таблиця 1

Порівняльна таблиця отриманих результатів моделей

Назва методу	Індекс GINI	Значення AUC	Точність моделі	Якість моделі
Бінарна логістична регресія	0,669	0,828	0,776	висока
Дерева рішень	0,612	0,766	0,754	прийнятна
Нечітка логіка	–	–	0,785	висока
Мережа Байєса	0,687	0,845	0,757	висока

Примітка: AUC (Area Under Curve) – обчислюється, наприклад, за методом трапецій:

$$AUC = \int f(x)dx = \sum_i \left[\frac{X_{i+1} + X_i}{2} \right] \cdot (Y_{i+1} - Y_i);$$

загальна точність моделі визначається відношенням правильно спрогнозованих випадків до їх загальної кількості.

Отримані результати свідчать, що найкращими моделями оцінки кредитоспроможності позичальників є ті, що побудовані за методами нечіткої логіки, логістичної регресії та мереж Байєса. Високі значення точності моделі також надає логістична регресія. Ці результати ще раз підтверджують доцільність використання нечіткої логіки, логістичної регресії та дерев рішень для оцінювання кредитоспроможності позичальника.

Метод на основі внутрішніх кредитних рейтингів. Метод на основі внутрішніх кредитних рейтингів (ВКР або Internal Rating Based Approach – IRB approach) – провідний метод оцінювання кредитних ризиків [3, 4]. Він надає змогу розробляти гнучкі механізми вимірювання очікуваних і неочікуваних втрат. За цим методом можна оцінювати індивідуальний і портфельний кредитний ризики. Основні показники, що характеризують обсяг потенційних втрат у разі використання ВКР, такі: 1) ймовірність дефолту позичальника (PD – probability of default), яка набуває значення від 0 до 1; 2) експозиція під ризиком (CE – credit exposure) – сума кредитної заборгованості; 3) покриття кредиту заставою (LGD – loss given default); набуває значення від 0 (кредит повністю покритий заставою) до 1 (кредит повністю не покритий заставою); 4) строк надання кредиту (M – maturity).

Всі дослідження під час оцінювання кредитного ризику ведуться саме в напрямі розроблення механізму розрахунку ймовірності дефолту позичальника. Для оцінювання ймовірності дефолту необхідно розробити механізм цього оцінювання. Постановку задачі оцінювання ймовірності дефолту в межах індивідуального кредитного ризику можна сформулювати так: на основі параметрів позичальника та значення кредиту x_i^j необхідно розробити процедуру оцінювання ймовірності дефолту PD_i :

$$PD_i = F(w^j, x_i^j), \quad (8)$$

де w^j – вагові коефіцієнти для параметрів x_i^j . Для розв'язання цієї задачі можна скористатись двома підходами: 1 – скоринговий підхід до побудови математичної моделі на основі статистики дефолтів за попередні періоди; 2 – експертний метод. Приклад застосування цього підходу до розрахунку обсягу очікуваних втрат для портфеля, який складається з 10 позичальників, ілюструє табл. 2.

Обсяг очікуваних кредитних втрат для цього портфеля становить 18,7 тис. грн. Отже, цей підхід надає змогу оцінювати обсяги можливих втрат для груп позичальників одночасно.

Таблиця 2

Приклад розрахунку очікуваних втрат для портфеля

Позичальник	CE, тис. грн.	Застава, тис. грн	LGD	PD позичаль- ника	ВКР позичаль- ника
1	220	200	0,05	0,0002	AAA
2	60	40	0,45	0,015	A
3	250	180	0,20	0,033	BB
4	120	20	0,95	0,045	B
5	50	0	0,90	0,022	A
6	250	100	0,45	0,018	BBB
7	140	80	0,85	0,0045	AA
8	200	20	0,88	0,017	AA
9	20	3,0	0,02	0,024	A
10	50	5,0	0,99	0,015	BBB

Примітка: AAA = 0,0001; AA = 0,005; A = 0,01; BBB = 0,02; BB = 0,03; B = 0,05; CCC = 0,1; CC = 0,25; C = 0,5; D = 1.

Висновки

Сучасні підходи до розв'язання задачі кредитування за умови мінімізації ризику можливих втрат потребують впровадження нових ефективних принципів управління ризиками та комп'ютерних систем підтримки прийняття рішень. Побудова таких систем передбачає розроблення та використання множини альтернативних методів аналізу даних, альтернативних моделей та відповідних критеріїв аналізу якості моделей і остаточного результату – ймовірності неповернення кредиту.

У роботі виконано аналіз деяких сучасних підходів до створення класифікаційних математичних моделей, сукупне використання яких надасть змогу приймати коректні обґрунтовані рішення стосовно видання кредитів клієнтам фінансової установи. Кращі результати класифікації клієнтів за фактичними статистичними даними отримано за нечіткою логікою, бінарними нелінійними моделями і мережею Байєса. Тобто моделі такого типу мають кращі показники за статистичними параметрами якості. Також перспективним є подальший розвиток методу на основі внутрішніх кредитних рейтингів, який забезпечує повнішу інформацію стосовно ситуації з кредитами. Наприклад, можна отримати оцінку можливих втрат.

У подальших дослідженнях необхідно удосконалювати вибрані типи класифікаційних моделей з метою підвищення якості класифікації клієнтів на дві групи. Також доцільно одночасно застосувати «ідеологічно» різні типи моделей – регресійні, ймовірнісні, нейромережеві та нейронечіткі.

1. Бідюк П.І. Моделі оцінювання ризиків кредитування фізичних осіб / П.І. Бідюк, С.О. Матрос // *Кібернетика та обчислювальна техніка*. – 2007. – № 153. – С. 87–95. 2. *Енциклопедія фінансового ризик-менеджмента* / под ред. О.О. Лобанова и А. В. Чугунова – М.: Альпина Паблишер, 2003. – 845 с. 3. Kiss F. *Credit scoring processes from a knowledge Management perspective* / F. Kiss // *Hungary Periodica Polytechnica*. – 2003. – vol. 11, № 1. – P. 95–110. 4. Jorion P. *Financial risk management handbook* / P. Jorion. – New Jersey: John Wiley & Sons, Inc., 2003. – 422 p. 5. Бідюк П.І. Оцінювання моделей стохастичної волатильності та VARUG на Java / П.І. Бідюк, М.М. Коновалюк // *Наукові праці: Комп'ютерні технології*. – Миколаїв: ЧДУ ім. Петра Могили, 2012. – Вип. 179, т. 191. – С. 14–20. 6. Бідюк П.І. Построение и методы обучения байесовских сетей / П.И. Бидюк, А.Н Терентьев, А.С. Гасанов // *Кібернетика и системный анализ*. – 2005. – № 4. – С. 134–147. 7. *Probabilistic networks and expert systems* / [Cowell R.G., Dawid A.Ph., Lauritzen S.L., Spiegelhalter D.J.]. – New York: Springer, 1999. – 323 p.