УДК 519.87

Івохін Є.В., д.ф.-м.н, доцент,
Присяженко О.Є., студент

Про один підхід до моделювання поведінкового скорингу

У статті розглянуто новий підхід до моделювання поведінки клієнта з використанням спеціальних множин простих чисел. Оцінки характеристик клієнтів задаються за допомогою пар простих чисел, банківські оцінки клієнтів визначаються цілими числами. Запропоновано метод розв'язання задачі побудови оцінки характеристик клієнта в будь-який момент часу на основі його попередньої поведінки, наведено приклад розв'язання задачі.

Ключові слова: поведінковий скоринг, скорингові моделі, спеціальні множини простих чисел.

Київський національний університет імені Тараса Шевченка, 03680, м.Київ, пр.Глушкова,4д
e-mail: ivohin@univ.kiev.ua, ollyajoke@gmail.com

Статтю представив д.т.н., проф. Волошин О.Ф.

Ivokhin E.V., Ph.D., associate professor,
Prysiazhenko O. I., student

On approach for modeling of behavioral scoring

In this paper a new approach to modeling a customer`s behavior on basis of special sets of prime numbers was discussed. The estimations of customers characteristics are given as pairs of prime numbers, bank estimations of customers is given as integers. The method of solution of the problem of customer`s assessment building in any subsequent moment on the basis of his previous behavior was suggested, an example of solving was given.

Key words: behavioral scoring, scoring models, special sets of prime numbers.

Taras Shevchenko National University of Kyiv, 03680, Kyiv, Gluskkva av., 4D

Аплікаційний скоринг [1], який на сьогоднішній день активно використовується для оцінки ризиків споживчих кредитів, являє собою поєднання двох основних характеристик стану клієнта: перша – це характеристика клієнта на основі даних заявки, друга – його кредитоспроможність в деякий наступний момент часу. Таким чином, цей феномен можна назвати статичним. Поведінковий скоринг, в свою чергу, є методом корекції оцінки ризиків споживчих кредитів з огляду на найостаннішу діяльність клієнта. Він замінює першу основну характеристику клієнта описом динаміки найостаннішої діяльності клієнта, але незмінною залишається друга характеристика.

Розглядаючи прибутковість клієнта для кредитора, потрібно використати інформацію про останню діяльність клієнта з метою оцінити його загальну подальшу поведінку, а не лише в деякий наступний момент часу. Для розробки оцінки прибутковості клієнтів необхідно оцінити майбутню діяльність споживача. Також потрібно спрогнозувати динаміку поведінки цієї оцінки або статусу заборгованості споживача, що також може бути способом оцінки загальної суми несплачених платежів даного портфелю клієнтів у кожний наступний період часу. Відповідні підрахунки необхідні для прогнозування резервів кредитора – суму коштів, яку кредитор повинен відкласти для покриття цих очікуваних грошових втрат. Така задача називається задачею формування резервів (забезпечення боргу). Найчастіше, для отримання цієї оцінки застосовуються ймовірнісні моделі Маркова. Підхід на основі ланцюгів Маркова не набув дуже широкої популярності у моделюванні динаміки поведінки клієнтів, але знайшов використання у знаходженні ймовірності дефолту, необхідної для розв’язання задачі формування резервів. З іншої сторони, для прогнозування неплатоспроможності (дефолту) клієнтів, можна використовувати моделі на базі ідей аналізу виживання (Survival analysis). Такі моделі також дозволяють оцінити прибутковість клієнтів в розрізі продукту, оскільки вони взаємодіють не лише з ризиком несплати платежів, а й з деякими іншими подіями, що впливають на прибуток, такими як дострокове погашення боргу. Дані підходи пов`язують останню динаміку змін поведінки клієнтів з динамікою зміни ймовірності несплати платежів протягом всього майбутнього періоду часу.

Таким чином, обидві статичні характеристики аплікаційного скорингу можуть бути трансформовані у динамічне представлення поведінки клієнта.

Динаміку змін оцінки поведінки клієнта можна формалізувати за допомогою спеціальних множин простих чисел [2], які можна використовувати при вирішенні багатьох прикладних задач в якості ефективних обчислювальних схем. Введемо деякі основні поняття та операції.

Означення 1. Множини невід'ємних простих чисел $P_j(a) \ge 0$, $j \in Z$, що належать інтервалу $[a, \infty)$ при $j \ge 0$ або інтервалу $[0, a)$ при $j < 0$ для заданого, необов'язково простого, цілого числа $a \ge 0$, будемо називати послідовностями простих чисел відносно числа a.

Розглянемо довільні прості числа $P_j(a) \ge 0$ з порядковими номерами $j \in Z$ з послідовності простих чисел відносно числа $a \ge 0$. Нескладно перевірити, що справедливі співвідношення:

1) $P_0(0) = 0$, $P_0(1) = 1$, $P_1(0) = 1$;

2) $P_0(a) = a$, якщо число $a \ge 0$ - просте, $P_0(a)$ - не існує, якщо $a \ge 0$ - непросте;

3) $P_j(a) \le P_k(a)$, якщо $j \le k$, $P_j(a) < P_k(a)$ при $j < k$, $j \in Z$, $k \in Z$;

4) $P_j(a) = P_j(a+1) = ... = P_j(a+l)$ для всіх $1 \le l < P_{j+1}(a) - P_j(a)$, $j = 0,1,2,...$, $a \ge 0$;

5) $P_j(a) = P_1(P_{j-1}(a)) =$
$= P_1(P_1(P_{j-2}(a))) = P_2(P_{j-2}(a)) = ... = P_{j-1}(P_1(a))$, якщо число $a \ge 0$ - просте, $j \in Z$;

6) $P_j(a) \ge a$ для усіх $j = 0,1,2,...$, якщо число $a \ge 0$ - просте, $P_j(a) > a$ для усіх $j = 1,2,...$, якщо $a \ge 0$ - непросте;

7) $P_j(a) < a$ для $j = -1, -2, ..., j_0$, де номер $j_0 < 0$ визначається як найменший індекс простого числа з послідовності, для якого $0 \le P_{j_0}(a) < a$.

Надалі будемо розглядати випадки, коли $a > 0$ - просте число або $a = 0$. У даному випадку існує число $P_0(a)$. Крім цього, виходячи з наведених вище властивостей, справедливі співвідношення $P_j(a)/P_k(a) \le 1$, і $P_j(a)/P_k(a) =$
$= P_j(a)/P_1(P_{k-1}(a)) = ... = P_j(a)/P_j(P_{k-j}(a)) =$
$= P_j(a)/P_{k-j}(P_j(a))$ для будь-яких $j, k \in Z$, $j \le k$.

Крім традиційних арифметичних операцій додавання, віднімання, множення та ділення, на послідовності простих чисел для довільного $P_j(a)$, $a \ge 0$, $j = 0,1,2,...$, введемо операцію зсуву на m, $m \in Z$, $j + m \ge j_0$, простих чисел у вигляді

$$P_j(a) \oplus m = P_{j+m}(a) = P_m(P_j(a)), \tag{1}$$

яка не повинна виводити за межі заданої послідовності ($j + m \ge j_0$), та операцію n-кратної композиції ($n \in Z$) відношення двох чисел $P_j(a)$ та $P_k(a)$, $j, k \in Z$, $j \le k$, у вигляді

$$\frac{P_j(a)}{P_k(a)} \circ n = \frac{P_j(a) \oplus n}{P_k(a) \oplus n} = \frac{P_{j+n}(a)}{P_{k+n}(a)} = \frac{P_n(P_j(a))}{P_n(P_k(a))}. \tag{2}$$

Лема 1. Довільне раціональне число $r = p/q$, $p \in Z$, $q \in N$, може бути представлене у вигляді

$$r = \prod_{i=1}^{s_p} P_{k_i}(a_{k_i}) \Big/ \prod_{j=1}^{s_q} P_{n_j}(a_{n_j}), \tag{3}$$

де s_p, s_q - кількості множників у представленні чисел $|p|$ та q у вигляді відповідних добутків елементарних дільників, $a_{k_i} \ge 0$, $a_{n_j} \ge 0$ - деякі цілі числа, $k_i, n_j \in Z$, $i = \overline{1, s_p}$, $j = \overline{1, s_q}$.

Доведення. Відомо, що будь-яке раціональне число r представляється у вигляді звичайного дробу $r = p/q$, $p \in Z$, $q \in N$.

Припустимо, що число r невід'ємне, $r \ge 0$. Якщо при цьому обидва числа p та q є простими числами, то маємо твердження леми. У даному випадку $s_p = 1, s_q = 1$, $P_{k_1}(a_{k_1}) = p$, $P_{n_1}(a_{n_1}) = q$ для деяких $k_1 \in Z$, $n_1 \in Z$ та $a_{k_1} \ge 0$, $a_{n_1} \ge 0$.

Якщо чисельник або знаменник або обидва числа p та q не є простими, то їх можна представити у вигляді добутків елементарних дільників, які є простими числами. При цьому, елементарні дільники записуються у добутках стільки разів, яка їх кратність.

Нехай s_p, s_q - кількості множників у представленні чисел p та q у вигляді добутків, де кожен з множників є простим числом з відповідних послідовностей простих чисел відносно чисел a_{k_i}, a_{n_j}, $k_i, n_j \in Z$, $i = \overline{1, s_p}$, $j = \overline{1, s_q}$. Представлення (3) для довільного раціонального числа r доведено.

У випадку, якщо число r від'ємне, то можна покласти $r = -r$ і усі викладки повторюються.

Наслідок 1. Довільне раціональне число $r = p/q$, $p \in Z$, $q \in N$, може бути представлене у вигляді

$$r = \prod_{i=1}^{s_p} P_{k_i}(a) \Big/ \prod_{j=1}^{s_q} P_{n_j}(a) , \qquad (4)$$

де s_p, s_q - кількості множників у представленні чисел $|p|$ та q у вигляді відповідних добутків елементарних дільників, $a \geq 0$ - деяке ціле число, $k_i \in Z$, $n_j \in Z$, $i = \overline{1, s_p}$, $j = \overline{1, s_q}$.

Наслідок 2. Довільне раціональне число $r = p/q$, $p \in Z$, $q \in N$, може бути представлене у вигляді

$$r = \prod_{i=1}^{s} P_{k_i}(1) \Big/ \prod_{j=1}^{s} P_{n_j}(1) , \qquad (4')$$

де s - максимальне значення з кількостей множників у представленні чисел $|p|$ та q у вигляді відповідних добутків елементарних дільників, $k_i, n_j \in N \cup \{0\}$, $i = \overline{1, s_p}$, $j = \overline{1, s_q}$.

Лема 2. Для будь-яких двох простих чисел $P_k(a)$, $P_n(a)$, $k \leq n$, $k, n \in Z$, з послідовності простих чисел відносно числа a та довільного числа $T \geq 0$ справедливе співвідношення:

$$(P_k(a)+T)/(P_n(a)+T) \geq P_k(a)/P_n(a) . \qquad (5)$$

Доведення. Нескладно перевірити, що для будь-яких $k \leq n$, $k, n \in Z$, маємо

$$(P_k(a)+T)/(P_n(a)+T) - P_k(a)/P_n(a) =$$
$$= T(P_n(a) - P_k(a))/((P_n(a)+T) * P_n(a)) .$$

Враховуючи властивість 3), отримуємо, що

$$(P_k(a)+T)/(P_n(a)+T) - P_k(a)/P_n(a) \geq 0 ,$$

звідки слідує нерівність (5). Лему доведено.

Розглянемо множину невід'ємних раціональних чисел $r(k,n)$, $k, n \in Z$, $0 \leq r(k,n) \leq 1$, які подаються у вигляді відношення двох простих чисел з послідовності відносно числа $a \geq 0$:

$$r(k,n) = P_k(a)/P_n(a) , \ k \leq n , \ k, n \in Z . \qquad (6)$$

Без обмежень загальності покладемо $a = 0$. При цьому, раціональні числа $r(k,n)$ будуть представлятися у вигляді

$$r(k,n) = P_k(0)/P_n(0) , \ k \leq n , \ k, n \in N \cup \{0\} . \qquad (7)$$

З урахуванням операції зсуву на m простих чисел у послідовності $P_j(0)$, $j \in Z$, маємо співвідношення

$$r(k+m, n+m) = r(k,n) \circ m , \ m \in N \cup \{0\} . \qquad (8)$$

Введемо до розгляду інші сукупності.

Означення 2. Множину раціональних чисел $r_T(k,n)$, $T \geq 0$, $k \leq n$, $k, n \in N \cup \{0\}$, вигляду

$$r_T(k,n) = (P_k(0)+T)/(P_n(0)+T) , \ k \leq n , \quad (9)$$
$$k, n \in N \cup \{0\} ,$$

назвемо T - послідовністю для множини чисел $r(k,n)$ і заданого $T \geq 0$.

Означення 3. Для заданої множини чисел $r(k,n)$, $k \leq n$, $k, n \in N \cup \{0\}$ множину чисел

$$r_*(k+m, n+m) = \frac{P_k(0) \oplus m_*}{P_n(0) \oplus m} , k \leq n , \ m \in Z , \quad (10)$$
$$k, n \in N \cup \{0\} ,$$

де m_*: $0 \leq m_* \leq m$ при $m \in N \cup \{0\}$, і $m_* \leq m$ при $m < 0$ - найбільше ціле число таке, що:

$$P_k(0) \oplus m \Big/ P_n(0) \oplus m \geq P_k(0) \oplus m_* \Big/ P_n(0) \oplus m ,$$

будемо називати нижньою спряженою множиною, а множину чисел

$$r^*(k+m, n+m) = \frac{P_k(0) \oplus m^*}{P_n(0) \oplus m} , \ k \leq n , \ m \in Z , \quad (11)$$
$$k, n \in N \cup \{0\} ,$$

де m_*: $m^* \geq m$ при $m \in N \cup \{0\}$ і $m \leq m^* \leq 0$ при $m < 0$ - найменше ціле число таке, що:

$$P_k(0) \oplus m \Big/ P_n(0) \oplus m \leq P_k(0) \oplus m^* \Big/ P_n(0) \oplus m ,$$

- верхньою спряженою множиною.

Лема 3. Справедливі співвідношення

$$r(k,n) \leq r_T(k,n) , \qquad (12)$$
$$r_*(k+m, n+m) \leq r(k+m, n+m) , \qquad (13)$$
$$r(k+m, n+m) \leq r^*(k+m, n+m) , \qquad (14)$$
$$r_*(k+m, n+m) \leq r_T(k+m, n+m) , \qquad (15)$$

для довільних $T \geq 0$, $k, n \in N \cup \{0\}$, $k \leq n$, $m \in Z$.

Доведення. Відповідно до леми 2, для довільного числа $T \geq 0$ справедлива нерівність $r_T(k,n) = (P_k(0)+T)/(P_n(0)+T) \geq P_k(0)/P_n(0) = r(k,n)$, і виконуються співвідношення: $r(k+m, n+m) =$

$$= \frac{P_k(0) \oplus m}{P_n(0) \oplus m} \geq \frac{P_k(0) \oplus m_*}{P_n(0) \oplus m} = r_*(k+m, n+m)$$

та $r(k+m, n+m) =$

$$= \frac{P_k(0) \oplus m}{P_n(0) \oplus m} \leq \frac{P_k(0) \oplus m}{P_n(0) \oplus m^*} = r^*(k+m, n+m) ,$$

для довільних $k, n \in N \cup \{0\}$, $k \leq n$, та будь-якого $m \in Z$. Таким чином, нерівності (12), (13), (14) доведено.

Нерівність (12) виконується для усіх

$k, n \in N \cup \{0\}$, $k \le n$. Тоді для будь-якого $m \in Z$ маємо $k + m \le n + m$, звідки для $\forall T \ge 0$ є справедливим співвідношення $r(k+m, n+m) \le$ $\le r_T(k+m, n+m)$. З урахуванням нерівності (13) остаточно отримуємо $r_*(k+m, n+m) \le$ $\le r_T(k+m, n+m)$, що відповідає нерівності (15).

Базуючись на наведених твердженнях, розглянемо модель функціонування поведінкового скорингу з застосуванням спеціальних множин простих чисел. Принцип оцінювання аналогічний до використання ланцюгів Маркова: маємо початковий стан клієнта $x^0 = (x_1^0, ..., x_N^0)^T$, де x_i^0, $i = \overline{1, N}$, – початкові значення характеристик, що є аплікаційними оцінками даних, отриманих із заявки на кредит. Кожен елемент вектора характеристик клієнта приймає значення з поля дійсних чисел і може бути представлений у вигляді відношення двох простих чисел $P_j^i(a) / P_k^i(a)$, $j, k \in Z$, $j \le k$, $i = \overline{1, N}$. Маємо деякі критерії f_i, $i = \overline{1, N}$, які визначаються банком для оцінювання клієнта і за допомогою яких формується вектор оцінки в деякий конкретний момент часу $f(t) = (f_1^t, ..., f_N^t)^T$, $t = 1,2,...$, елементи якого цілими числами. Для визначення характеристик клієнта в наступний момент часу, необхідно використати операцію n-кратної композиції відношення двох простих чисел з $n_i = f_i^t$, $i = \overline{1, N}$, $t = 1,2,...$. В результаті застосування даної операції отримуємо вектор $x^t = (x_1^t, ..., x_N^t)^T$, $t = 1,2,...$, кожен елемент якого визначається наступним чином:

$$x_i^t = x_i^{t-1} \circ f_i^t = P_j^i(a) / P_k^i(a) \circ f_i^t, \; i = \overline{1, N}.$$

При застосуванні даної операції важливо, щоб зберігалася тенденція до збільшення (зменшення) оцінок характеристик клієнта, якщо відповідна оцінка банку додатна (від'ємна). Отже, якщо $f_i^t \ge 0$, $t = 1,2,...$, $i = \overline{1, N}$, то в якості наступного значення доцільно обрати елемент з верхньої спряженої множини чисел, а якщо ж $f_i^t < 0$ – з нижньої спряженої множини чисел.

Розглянемо приклад. Оберемо 5 критеріїв:
1) оцінка показника кредитної історії клієнта;
2) оцінка показника матеріального стану клієнта;
3) оцінка терміну співпраці з банком;
4) оцінка статусу проживання клієнта;
5) оцінка рівня відповідності банківського продукту кредитній історії клієнта.

Оцінка стану клієнта залежить від показників по цим критеріям в кожний деякий момент часу та змінюється під впливом своїх складових. Наприклад, якщо клієнт не виконав свої зобов'язання по кредиту у визначений термін та пропустив платіж, то, відповідно, загальна кількість прострочених платежів зросте, що негативно вплине на його оцінку. Також ступінь довіри банка до клієнта, з яким він давно співпрацює, буде вищим, ніж до людини, що звернулася вперше, за умови рівності решти критеріїв.

Припустимо, що

$$x^0 = (3/23, 7/13, 5/7, 11/13, 3/5)^T,$$

$$f(1) = (5, 4, -3, 2, -1)^T.$$

Тоді

$$x^1 = (3/23, 7/13, 5/7, 11/13, 3/5)^T \circ (5, 4, -3, 2, -1)^T =$$
$$= (P_5(3)/P_5(23), P_4(7)/P_4(13), P_{-3}(5)/P_{-3}(7),$$
$$P_2(11)/P_2(13), P_{-1}(3)/P_{-1}(5))^T =$$
$$= (17/43, 19/29, 1/2, 17/19, 2/3)^T,$$

а з урахуванням вищенаведеного зауваження

$$x_5^1 = \frac{P_0(3) \oplus (-1)_*}{P_0(5) \oplus (-1)} = \frac{P_{-2}(3)}{P_{-1}(5)} = \frac{1}{3}.$$

Остаточно отримуємо

$$x^1 = (17/43, 19/29, 1/2, 17/19, 1/3)^T.$$

Продовжуючи цей процес далі, можна ефективно оцінювати стан характеристик клієнта в будь-який момент часу, що дозволяє мінімізувати ризики при наданні споживчих кредитів. З урахуванням *Леми 3* за сприятливої співпраці клієнта з банком, позитивній кредитній історії та особистих матеріальних показниках ймовірність того, що надання кредиту буде вигідним для банку, зростає, а негативні фактори поведінки клієнта зменшують цю ймовірністю, Використання цілих чисел у розрахунках може спростити процес оцінювання та інтерпретацію вихідних даних на практиці, а вихідна інформація забезпечує підтримку прийняття остаточного рішення щодо надання кредиту певній особі на базі скорингової моделі.

Список використаних джерел

1. *Thomas L. C., Ho J., Scherer W. T.* Time will tell: Behavioural scoring and the dynamics of consumer credit assessment, 2012. – 33р.
2. *Ivokhin E.V., Vadniov D.O.* Use of special sets of prime numbers in knapsack method encryption// Proc. VI International Liashko conf. "Comput. and Appl. Math." – Kyiv, 2013. – P.126-127.

Надійшла до редколегії 25.11.2013