УДК 336.77

Г. О. Зінов'єва,
студентка, Факультет менеджменту та маркетингу,
Національній технічний університет України "КПІ"

ІНТЕРВАЛЬНА МОДЕЛЬ ПРИЙНЯТТЯ ФІНАНСОВИХ РІШЕНЬ

У даній роботі розглядається проблематика прийняття кредитних рішень, а саме — технології скорингу. Основна мета дослідження полягає у виявленні неблагонадійних потенційних позичальників, створенні інтервальної моделі прийняття фінансових рішень, аналізі кредитоспроможності клієнтів банку та окресленні потенційних напрямів подальшого розвитку та удосконалення роботи за даною тематикою.

У даній роботі були побудовані скорингові моделі на основі логістичної регресії та методом k-найближчого сусіда, був проведений їх ROC-аналіз, що показав відмінну ефективність алгоритмів. На основі цього було побудовано модель прийняття колективного рішення з урахуванням неоднозначної відповіді про надання кредиту згаданих раніше моделей. Узагальнюючи роботу, була сформульована інтервальна модель прийняття кредитних рішень.

In this work an author is examine problematika of acceptance of credit decisions, namely technologies of skoring. The primary purpose of research consists exposure of unreliable potential borrowers, creation of mathematical model of acceptance of credit decisions, analysis of solvency of clients, Diamont bank and lineation of potential directions of subsequent development and improvement of work after this subject.

In this work skoringovi models were built on the basis of logistic regression and by the k — neighbors method, was conducted them ROC is an analysis which rotined a next-door neighbour excellent efficiency of algorithms. On the basis of it, the model of acceptance of collective decision was built taking into account an ambiguous answer about a grant the credit of mentioned ranishe models. Summarizing the work was formulated interval model make credit decisions.

Ключові слова: скоринг, ROC-аналіз, скорингова модель, кредитоспроможність, погашення, модель прийняття остаточного кредитного рішення, інтервальна модель прийняття рішень.

ВСТУП

Скоринг являє собою математичну або статистичну модель, за допомогою якої на основі кредитної історії "минулих" клієнтів банк намагається визначити, наскільки велика ймовірність, що конкретний потенційний позичальник поверне кредит у строк.

У даній статті проаналізовано підходи до моделювання скорингових систем та їх оцінки методом ROC-аналізу. У зв'язку з тим, що розвиток системи кредитування в Україні стрімко зростає, моделювання цього процесу є важливим і актуальним питанням.

ПОСТАНОВКА ЗАДАЧІ

Поставимо моделі задачу виявлення неблагонадійних потенційних позичальників. Але, оскільки в скорингу загальноприйняте, що чим вище рейтинг клієнта, тим вище його кредитоспроможність, то будемо вважати позитивним результатом успішне погашення позики, а негативним — дефолт по кредиту.

Тоді, проектуючи за цих умов визначення чутливості й специфічності на скоринг, можна вважати, що скорингова модель із високою специфічністю відповідає консервативній кредитній політиці (частіше відбувається відмова у видачі кредиту), а з високою чутливістю — політиці ризикованих кредитів. У першому випадку мінімізується кредитний ризик, пов'язаний із втратами позички, відсотків і додаткових витрат на повернення кредиту, а в другому — комерційний ризик, пов'язаний з упущеною вигодою.

Об'єкт моделювання в роботі — це скорингові системи, що визначають надійність позичальника. Їх основною метою є визначення некредитоспроможних клієнтів.

МЕТОДОЛОГІЯ

У роботі проаналізовані та використані математичні методи для вирішення поставленої задачі.

Розглянувши приклади моделей прийняття рішень і основні методи, які забезпечують підтримку кредиторам при виборі рішень по споживчому кредиту, для роботи було обрано побудову скорингових моделей на основі логістичної регресії та методом k-найближчого сусіда, які відповідають практиці комерційного банку "Діамант".

Логістична регресія є статистичним інструментом для розрахунку коефіцієнтів (балів) скорингової карти на основі накопиченої кредитної історії.

Для побудови рівняння регресії (1) використовуються анкетні дані позичальників: вік, стать, знаходиться у шлюбі (так/ні), кількість утриманців, сукупний дохід, досвід роботи, період проживання в регіоні, ринкова вартість власної нерухомості, щомісячний платіж за кредитом.

Таким чином, рівняння регресії матиме наступний вигляд:

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (1).$$

Метод "найближчого сусіда" ("nearest neighbour") відноситься до класу методів, робота яких грунтується на зберіганні даних в пам'яті для порівняння з новими елементами. При появі нового запису для прогнозування знаходяться відхилення між цим записом і подібними наборами даних, найбільш подібна (або ближній сусід) ідентифікується.

При розгляді нового клієнта банку його атрибути порівнюються зі всіма існуючими клієнтами даного банку (дохід, вік і т.д.). Безліч "найближчих сусідів" потенційного клієнта банку вибирається на підставі найближчого значення доходу, віку і т.д.

Приклад реалізації даної моделі зображен на рис. 1 (STATISTICA 6).

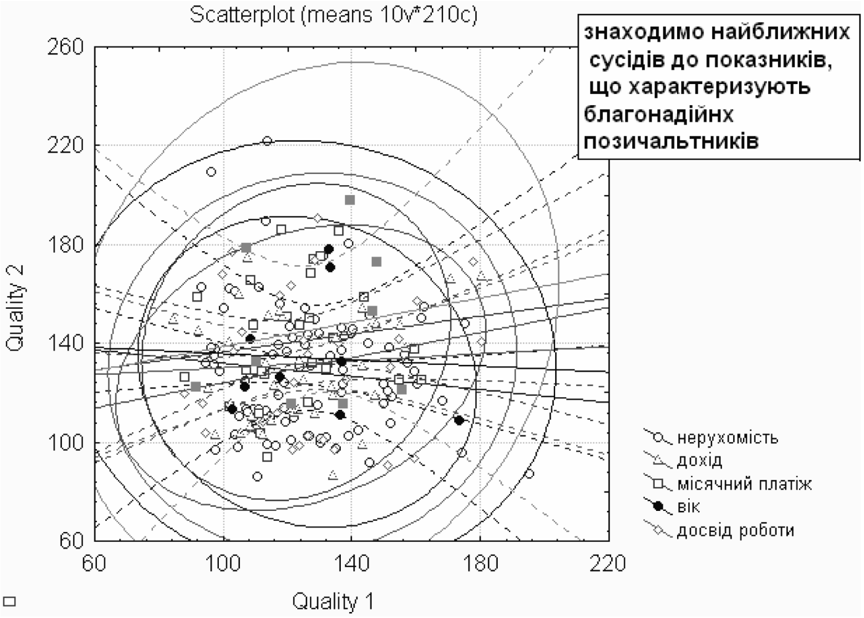

Рис. 1. Моделі на основі методу k-найближчих сусідів

Проведемо оцінку якості алгоритмів класифікації методом ROC-аналізу. ROC-крива показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих негативних прикладів. Позитивний результатом є успішне погашення позики, а негативним — дефолт по кредиту.

ROC-крива виходить у такий спосіб:

1. Для кожного значення порога відсікання, що міняється від 0 до 1 із кроком *dx* (наприклад, 0.01), розраховуються значення чутливості *Se* і специфічності *Sp*. Як альтернатива, порогом може бути кожне наступне значення приклада у вибірці.

2. Будується графік залежності: по осі *Y* відкладається чутливість (частка істинно позитивних випадків), по осі *X* — *100%-Sp* $Se = TPR = \dfrac{TP}{TN + FN}100\%$ (сто відсотків мінус специфічність), $Sp = \dfrac{TN}{TN + FP}100\%$ (частка істинно негативних випадків).

— TP (True Positives) — істинно позитивні випадки;

— TN (True Negatives) — істинно негативні випадки;

— FN (False Negatives) — позитивні приклади, класифіковані як негативні (помилка I роду).

— FP (False Positives) — негативні приклади, класифіковані як позитивні.

Оцінка моделі отримується обчисленням площі AUC (Area Under Curve) під кривою.

Таблиця 1. Експертна шкала показника площі під кривою

Інтервал AUC	Якість моделі
0.9-1.0	Відмінне
0.8-0.9	Дуже гарне
0.7-0.8	Гарне
0.6-0.7	Середнє
0.5-0.6	Незадовільне

$$AUC = \int f(x)dx = \frac{1}{2}\sum_i (x_{i+1} + x_i)(y_{i+1} + y_i) \quad (2).$$

Ідеальна модель володіє 100% чутливістю й специфічністю.

Спираючись на політику банку: скорингова модель із високою специфічністю відповідає консервативній кредитній політиці (частіше відбувається відмова у видачі кредиту), а з високою чутливістю — політиці ризикованих кредитів.

Вимога балансу між чутливістю й специфічністю, тобто коли:

$$Cut\_off_0 = \min_k |Se_k - Sp_k| \quad (3).$$

За даними таблиць отриманих значень чутливості і специфічності, були побудовані та порівняні ROC криві (рис. 4, 5) для скорингових моделей на основі логістичної регресії та на основі методу k-найближчого сусіда. Опираючись на результати виконаної роботи, дані моделі можна охарактеризувати як відмінні (за експертною шкалою якості моделі). Це свідчить про те, що банк надає кредити благонадійним позичальникам, у більшості випадків керуючись даними кредитно-скорингової програми.

На самому початку дослідження перед нами була поставлена задача побудувати модель прийняття остаточного кредитного рішення, яка б з легкістю виявляла неблагонадійних позичальників.

Нехай два алгоритми (*N=2*) незалежно один від одного приймають особисті рішення про надання або ненадання кредиту, який може знаходитись в одному з двох можливих станів (*M=2*) з апріорними ймовірностями:

$$P(V_1) \text{ і } P(V_2) = 1 - P(V_1).$$

Нехай відомі ймовірності помилок першого та другого роду особистих рішень незалежних експертів

$$P^{(i)}(E \mid V_k) = P(\delta_i = m \mid V_k), \; m, k, i = 1,2, \; k \neq m \quad (4).$$

Тоді колективне рішення *D=m, m=1, 2* є оптимальним з точки зору мінімуму середньої ймовірності помилки на множині можливих ситуацій, якщо в ситуації $S_{m_1 m_2}$, $m_1 \neq m_2$ колективне рішення $D \neq 1$ приймається на користь V_1 за умови:

$$P(V_1)P^{(m_2)}(E \mid V_1)[1 - P^{(m_1)}(E \mid V_1)] > P(V_2)P^{(m_1)}(E \mid V_2)[1 - P^{(m_2)}(E \mid V_2)] \quad (5),$$

і рішення $D=2$ на користь V_2 при

$$P(V_1)P^{(m_2)}(E \mid V_1)[1 - P^{(m_1)}(E \mid V_1)] < P(V_2)P^{(m_1)}(E \mid V_2)[1 - P^{(m_2)}(E \mid V_2)] \quad (6).$$

Припускається умовна незалежність ймовірностей помилок експертів:

$$P^{(m_1 m_2)}(E \mid V_k) = P^{(m_1)}(E \mid V_k) P^{(m_2)}(E \mid V_k),$$

де $P^{(m_1 m_2)}(E \mid V_k)$ — ймовірність помилки обох експертів за умови знаходження об'єкта в k-му стані.

Для подальшого використання моделі (5), (6) введемо позначення:

$$L = P(V_1)P^{(m_2)}(E \mid V_1)[1 - P^{(m_1)}(E \mid V_1)] \quad (7),$$

$$W = P(V_2)P^{(m_1)}(E|V_2)[1 - P^{(m_2)}(E|V_2)] \quad (8).$$

Таким чином, приймається рішення на користь першого алгоритму, якщо $L > W$, на користь другого, якщо $L < W$.

Поставимо нову задачу: побудувати за неповними апріорними даними інтервальну модель, що в умовах суперечності особистих рішень алгоритмів з довірчою ймовірністю буде оптимальною з точки зору критерію мінімуму середньої ймовірності помилки колективного рішення.

Зрозуміло, що заміна ймовірностей $P^{(i)}$, i=1, 2, що були застосовані в моделі (5), (6), точковими оцінками можлива лише за великої кількості експериментів n. Тому розглядаємо можливість узагальнення моделі (5), (6) на випадок, коли замість точкових значень ймовірностей $P^{(i)}$ використовуються довірчі інтервали I_i.

Нехай тепер замість точкових значень ймовірностей $P^{(i)}$ маємо лише інформацію про частоти $P^{*(i)}$, i=1, 2 помилок, що припустилися алгоритм на вибірці обмеженого обсягу з відомими станами об'єкта.

Частота $P^{*(i)}$ випадкової події, що обчислена на виборці обсягом n, з довірчою імовірністю β визначає довірчий інтервал I_i для ймовірності P, який можна записати у формі центр-радіус:

$$I_i = \langle P^{c^{(i)}}, r^{(i)} \rangle \quad (9),$$

$$I_i = \langle P^{\min^{(i)}}(n,\beta), P^{\max^{(i)}}(n,\beta) \rangle \quad (10),$$

де

$$P^{\max^{(i)}} = \frac{P^{*(i)} + \frac{t_\beta^2}{2n} - t_\beta \sqrt{\frac{P^{*(i)}(1-P^{*(i)})}{n} + \frac{t_\beta^2}{4n^2}}}{1 + \frac{t_\beta^2}{n}} \quad (11),$$

$$P^{\min^{(i)}} = \frac{P^{*(i)} + \frac{t_\beta^2}{2n} + t_\beta \sqrt{\frac{P^{*(i)}(1-P^{*(i)})}{n} + \frac{t_\beta^2}{4n^2}}}{1 + \frac{t_\beta^2}{n}} \quad (12).$$

У цих співвідношеннях:

$$t_\beta = \arg\Phi^*\left(\frac{1+\beta}{2}\right),$$

де $\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-\tau^2}{2}} d\tau$ — нормальна функція розподілу.

Таким чином:

$$P^{c^{(i)}} = \frac{P^{*(i)} + \frac{t_\beta^2}{2n}}{1 + \frac{t_\beta^2}{n}}, \text{ а } r^{(i)} = \frac{t_\beta \sqrt{\frac{P^{*(i)}(1-P^{*(i)})}{n} + \frac{t_\beta^2}{4n^2}}}{1 + \frac{t_\beta^2}{n}} \quad (13).$$

Для отримання інтервального аналогу L, W (5), (6) точкових значень L, W змінимо всі операції і операнди в аналітичному вигляді для вираження L, W на інтервальні. Оскільки $P(V_1) \in I_{v_1}$, а $P(V_2) \in I_{v_2}$, то $P(V_2)[1 - P(V_1)] \in I_{v_2}(1 - I_{v_1}) = L$, $\lambda P(V_1)[1 - P(V_2)] \in \lambda I_{v_1}(1 - I_{v_2}) = W, \lambda = const$,

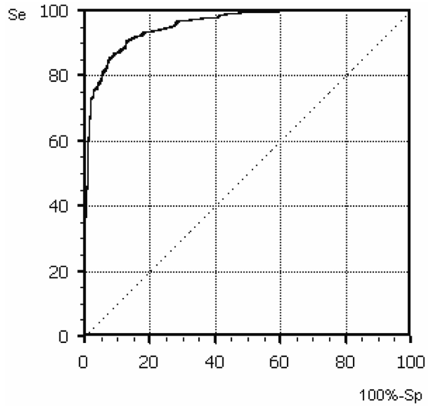

Рис. 2. Графік залежності специфічності і чутливості

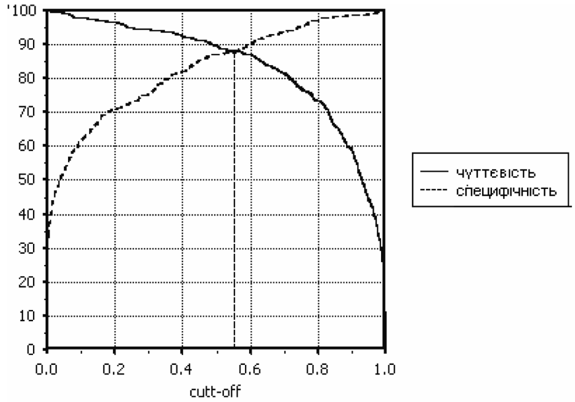

Рис. 3. Поріг відсікання за перетину графіків чутливості

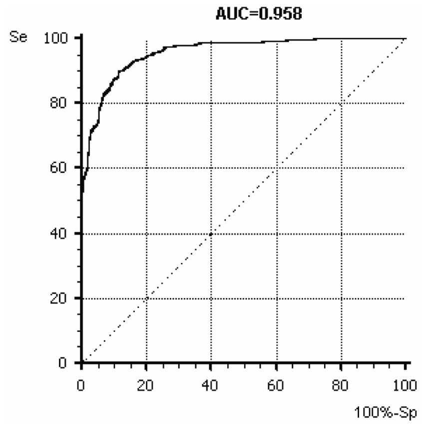

Рис. 4. ROC-крива для моделі на основі логістичної регресії

то маємо:

$$L = \langle P^c_{v_2}, r_{v_2} \rangle (1 - \langle P^c_{v_1}, r_{v_1} \rangle) = \langle P^c_{v_2}, r_{v_2} \rangle (1 - P^c_{v_1}, r_{v_1}) =$$
$$= \langle P^c_{v_2}(1 - P^c_{v_1}) + r_{v_2} r_{v_1}, P^c_{v_2} r_{v_1} + r_{v_2}(1 - P^c_{v_1}) \rangle \quad (14),$$

$$W = \lambda \langle P^c_{v_1}, r_{v_1} \rangle (1 - \langle P^c_{v_2}, r_{v_2} \rangle) = \lambda \langle P^c_{v_1}, r_{v_1} \rangle (1 - P^c_{v_2}, r_{v_2}) =$$
$$= \lambda \langle P^c_{v_1}(1 - P^c_{v_2}) + r_{v_2} r_{v_1}, P^c_{v_1} r_{v_2} + r_{v_1}(1 - P^c_{v_2}) \rangle \quad (15).$$

Будь-яке значення з інтервалу L буде більшим (чи меншим) будь-якого значення з інтервалу W, якщо $L = \langle L^c, r_L \rangle$, $W = \langle W^c, r_W \rangle$ не перетинаються, тобто:

$$L^c - r_L > W^c + r_W \quad (16)$$

чи

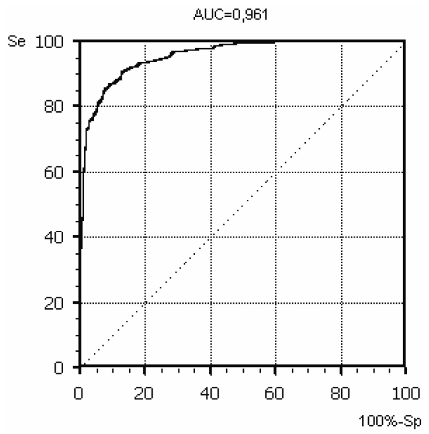

Рис. 5. ROC-крива для моделі методом k-найближчого сусіда

$$L^c + r_L < W^c - r_W \quad (17).$$

З (14) та (16) слідує:

$$\langle P^c{}_{v_2}(1-P^c{}_{v_1}) + r_{v_2} r_{v_1}, P^c{}_{v_2} r_{v_1} + r_{v_2}(1-P^c{}_{v_1})\rangle >$$
$$> \lambda\langle P^c{}_{v_1}(1-P^c{}_{v_2}) + r_{v_2} r_{v_1}, P^c{}_{v_1} r_{v_2} + r_{v_1}(1-P^c{}_{v_2})\rangle$$

Звідси за допомогою перетворень маємо:

$$(P^c{}_{v_2} - r_{v_2})(1-P^c{}_{v_1} - r_{v_1}) > \lambda(P^c{}_{v_1} + r_{v_1})(1-P^c{}_{v_2} + r_{v_2}) \quad (18).$$

Аналогічно з (15) та (17) маємо:

$$(P^c{}_{v_2} + r_{v_2})(1-P^c{}_{v_1} + r_{v_1}) < \lambda(P^c{}_{v_1} - r_{v_1})(1-P^c{}_{v_2} - r_{v_2}) \quad (19).$$

Таким чином, отримані умови (18) та (19), в яких замість значень ймовірностей (7), (8) $P_{v_1}^{(i)}, i = 1,2$ використовуються частоти алгоритмів та радіуси довірчих інтервалів.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Були побудовані та порівняні дві моделі: модель на основі логістичної регресії, модель на основі методу k-найближчого сусіда. Більш ефективною виявилась побудована модель на основі логістичної регресії. Для прийняття остаточного кредитного рішення була побудована модель прийняття кредитного рішення. Для перевірки даних алогитмів вони були застосовані на практиці для клієнтів комерційного банку. Узагальнюючи роботу, була побудована інтервальна модель прийняття рішень, що в умовах суперечності особистих рішень алгоритмів з довірчою ймовірністю буде оптимальною з точки зору критерію мінімуму середньої ймовірності помилки колективного рішення.

ВИСНОВОК

Після того як алгоритм кредитного скоринга розроблений, він повинен бути вбудований у фронтальну систему. Така фронтальна система може використовуватися для підтримки прийняття рішень кредитним офіцером, коли він розглядає заявки аплікантів на отримання кредиту, або для самодіагностики аплікантів за допомогою web-сервісу, представленого на сайті банку в Інтернеті.

У фронтальній системі вирішується досить просте завдання: за інформаційними ознаками апліканта визначити його скоринг-бал і порівняти його з балом відсікання.

Оскільки сучасні інформаційні технології розробки скорингових алгоритмів, як правило, передбачають генерацію програмного коду отриманого скорингового алгоритму, практична реалізація створення фронтального додатка праці не представляє.

Із часом предиктивна потужність скорингового алгоритму внаслідок об'єктивних причин знижується. Наступає час для повторного розрахунку скорингових алгоритмів і заміни у фронтальних додатках колишньої процедури скорингу новою. Вибір моменту перерахунку скорингової моделі є самостійним завданням, що може бути вирішеним на основі відповідних статистичних критеріїв. Однак якщо в банку на ділянку скоринг-аналитика виділений фахівець на постійній основі, то перерахунок скорингових алгоритмів (принаймні теоретично) може здійснюватися з будь-якою періодичністю в міру поповнення бази даних новими кредитними історіями.

В Україні впровадження скоринга повинне здійснюватися поступово. Для початку можна зробити автоматизовану експертну систему попередньої оцінки позичальників, що буде автоматично відсівати свідомо "погані" ризики, а на розгляд кредитного комітету пропонувати ризики "гарні" і "прикордонні". Дана система дозволить банківським працівникам швидко ухвалювати рішення щодо кредитування, регулювати обсяги кредитування залежно від ситуації на ринку й визначати оптимальне співвідношення між прибутковістю кредитних операцій і рівнем ризику.

Література:

1. Черкашенко, В. Ріст споживчого кредитування, кризи й скоринг / В. Черкашенко // Фінансова консультація. — 2005. — №1—№2.
2. Davis J., Goadrich M. The Relationship Between Precision-Recall and ROC Curves / J. Davis, M. Goadrich // Proc. Of 23 International Conference on Machine Learning. — Pittsburgh, PA, 2006.
3. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers / T. Fawcett // 2004 Kluwer Academic Publishers.
4. Churchill G.A., Nevin J.R., Watson R.R. / G.A. Churchill, J.R. Nevin, R.R. Watson // The role of credit scoring in the loan decision. Credit World. — March, 1977; Myers J.H., Forgy E.W. The development of numerical credit evaluation systems // Journal of American Statistical Association. — September, 1963. 15000 p.
5. Багатомірний статистичний аналіз в економіці /За ред. проф. В.Н.Тамашевича. — М.: ЮНИТИ-ДАНА, 1999. — 598 с.
6. Непман Дж. Математическое понятие решения Теория игр и математическое поведение / Дж. Непман. О. Моргенштерн. М.: Наука, 1971. 435 с.
7. Барабаш Ю.Л. Коллективные статистические решения при распознавании. — М: Радио и связь, 1983. — 224 с.
8. Айзерман М.А. Выбор вариантов: основы теории: монография / М.А. Айзерман,Ф.Т.Алескеров. — М.: Наука, 1990. — 236 с.
9. Бейко И.В. Задачи принятия решений в конфликтной ситуации. — Киев: Наукова думка, 1994. — 137 с.
10. Вентцель Е.С. Теория вероятностей. — М.: Наука, 1969. — 576 с.
11. Жуковская О.А. Интервальные модели принятия коллективных решений в конфликтных ситуациях // Международная конференция "Проблемы управления и приложения". — Минск, 2005. — 32 с.
12. Жуковская О.А., Файнзильберг Л.С. Интервальное обобщение байесовской модели принятия коллективного решения в конфликтных ситуациях // Кибернетика и системный анализ. — 2005. — №3. — С. 133—144.
13. Жуковская О.А., Файнзильберг Л.С. Формальная оценка квалификации эксперта на основе байесовской модели и методов интервального анализа // Проблемы управления и информатики. — 2005. — №3. — С. 103—115.

Стаття надійшла до редакції 10.02.2010 р.