

Моделювання в економіці, організація виробництва та управління проектами

УДК 004.932.2

В.О. Гороховатський, А.М. Кобилін, Р.В. Семенцов

*Харківський інститут банківської справи університету банківської справи
Національного банку України, Харків*

ЗАСТОСУВАННЯ ІНТЕЛЕКТУАЛЬНИХ ТЕХНОЛОГІЙ АНАЛІЗУ ДАНИХ ДЛЯ ОБРОБЛЕННЯ ІНФОРМАЦІЇ ПРО ДІЯЛЬНІСТЬ БАНКІВ

Обговорюється проблема оцінювання ефективності роботи банків на основі аналізу значень системи показників. Застосування методів кластерного аналізу векторних даних про рівень капіталізації дає можливість виділити групи банків з метою виявлення загальних тенденцій їх розвитку та подальшого управління процесом їх функціонування. Наведено результати аналізу даних про діяльність банків Харківського регіону.

Ключові слова: *аналіз діяльності банків, система показників, кластеризація, показники капіталізації, нормалізація компонентів, прийняття рішень.*

Вступ

Сучасний недостатній рівень банківського менеджменту часто приводить до погіршення надійності банківської діяльності і навіть до банкрутства деяких вітчизняних банків.

Необхідність оцінювання діяльності банку до початку співробітництва з ним обумовлена його роллю фінансового посередника. В умовах жорсткої конкуренції менеджмент банків зобов'язаний адекватно реагувати на зміни середовища задля забезпечення стійкості кожного банківського закладу окремо і банківської системи в цілому. Від уміння швидко реагувати на зміни, що впливають на стабільність банку, залежить успіх в управлінні його діяльністю. Банківська стабільність означає постійну здатність банку відповідати за своїми зобов'язаннями і забезпечувати прибутковість на рівні, достатньому для нормального функціонування у конкурентному середовищі. До основних критеріїв функціонування банку належить показник капіталізації, зменшення якого нижче заданого рівня означає підвищення ризикованості у діяльності банку [1, 2]. Обсяг капіталізації вітчизняної банківської системи на цей час часто не відповідає економічним потребам суб'єктів господарювання та вимогам зарубіжних інвесторів. Значного поширення у банківстві набуло застосування комп'ютерних інформаційних технологій, що наряду з автоматизацією процесів у банківському середовищі переслідує мету моніторингу діяльності банку за визначеною системою економічних показників [3]. Наприклад, можна виконати кластеризацію числових векторів, що містять

показники діяльності, з метою групування банків за близькими результатами. Отримані кластери можуть слугувати основою, наприклад, для прийняття рішення про загальні тенденції розвитку банків у рамках кластерів або необхідність додаткової капіталізації окремих груп банків. У цілому детальний аналіз груп банків за галузевими чи універсальними системами показників їх діяльності сприяє виробленню моделей і програм, що націлені на покращення показників функціонування банків.

Мета роботи – застосування методів кластерного аналізу даних при обробленні економічних показників діяльності банків для прийняття ефективних менеджерських рішень.

Задачі дослідження – групування економічних об'єктів-банків за системою показників, що виявляє скриті закономірності та тенденції їх діяльності, а також перевірка та порівняння різних кластерних методів на конкретній множині економічних даних.

Основні показники капіталізації банків

При дослідженні капіталізації банків використовують в основному такі показники: мультиплікатор капіталу (МК), ІСВК (індекс якісного складу власного капіталу), прибутковість власного капіталу (ПВК) [4]. Їх обчислення спирається на величину сукупних активів (СА), статутного капіталу (СК), чистого прибутку/збитку (ЧПЗ), власного капіталу (ВК).

Мультиплікатор капіталу відображає рівень ефективності управління ресурсами і розраховується як:

$$МК = СА/ВК. \quad (1)$$

Індекс якісного складу власного капіталу показує залежність банку від його засновників (акціонерів, учасників):

$$ICBK=CK/BK. \quad (2)$$

Прибутковість власного капіталу показує, скільки грошових одиниць чистого прибутку приносить одиниця власного капіталу:

$$PK=ЧПЗ/BK. \quad (3)$$

Як бачимо, ключовим показником для всіх критеріїв (1) – (3) є власний капітал BK. У результаті маємо числовий вектор K показників капіталізації

$$K=(MK, ICBK, PK). \quad (4)$$

Із літератури [2, 4] відомо, що оптимальні значення таких показників, як мультиплікатор капіталу МК знаходяться у межах 10 – 20, а показника ICBK – від 4 до 10.

Кластеризація даних: визначення, параметри, методи

Кластеризація є однією з найбільш важливих завдань аналізу даних (сучасна назва – Data Mining). Кластеризація – це розбиття певної множини об'єктів на підмножини (кластери), що не перетинаються, таким чином, щоб кожен кластер містив схожі об'єкти, а об'єкти різних кластерів відрізнялися між собою [5, 6]. Кластеризацію використовують, зокрема, в задачах статистичного аналізу даних, класифікації, розпізнаванні образів тощо [6].

Кластерний аналіз (англ. cluster analysis) – багатовимірна статистична процедура, що впорядковує вибірку об'єктів в однорідні групи, тобто отримує множину кластерів. Задача кластеризації відноситься до сфери інтелектуальної статистичної обробки, а також до широкого класу задач навчання без учителя. Основними цілями кластеризації вважаються: розуміння даних з подальшим спрощенням процедур оброблення; стиснення даних та виявлення нових нетипових об'єктів. На даний момент найбільш поширеними вважаються близько 50 методів кластеризації.

Стосовно об'єктів-банків задача кластеризації зводиться до виявлення груп банків з близькими показниками діяльності, наприклад, у відповідності до критерію близькості трьохвимірних векторів виду (4). Різниця між сформованими групами банків чи інших об'єктів, як правило, вважається суттєвою для дослідження або для прийняття рішення. У статті [4] розбиття множини банків на підставі експертної кластеризації на три кластери дало можливість виробити рекомендації щодо їх ефективності та надійності в умовах кризи.

Реалізація кластерних процедур, як правило, пов'язана з рядом параметрів [5], найбільш значущими серед яких вважають: застосовувану міру схожості між елементами та кластерами, значення

порогу еквівалентності для цієї міри, на основі якого реалізується віднесення або не віднесення об'єкта до кластеру. Важливими також є кількість кластерів та пов'язаний з конкретним підходом спосіб визначення і безпосередньо значення критерію, що оцінює якість (ефективність) кластеризації.

У загальному випадку застосування різних підходів кластеризації до одних і тих же даних дає, як правило, різні результати у вигляді множини кластерів. Методи, що показують кращу ефективність, у більшості випадків виявляються менш швидкодіючими.

Зрозуміло, що результат кластеризації повністю визначається вхідними даними, що обробляються. З цієї причини дуже важливим є попередній аналіз та оброблення даних, наприклад виконання нормування вхідних даних до однакового діапазону. У протилежному випадку дані з великими значеннями можуть повністю викривити або знизити вплив решти значень.

Розглянемо принципи побудови двох досить поширених методів кластеризації: метод k-середніх та EM-метод, які реалізовано у програмному продукті STATISTICA [7].

Метод k-means – це ітеративна процедура [5], яка ділить задану сукупність елементів на k кластерів, точки яких є максимально наближеними до їх центрів, а сама кластеризація відбувається за рахунок зміщення цих центрів. Метод прагне мінімізувати сумарне квадратичне відхилення точок кластерів від центрів цих кластерів (критерій ефективності). Головні переваги методу k-середніх – його простота та швидкість виконання. Метод k-середніх досить зручний для кластеризації значної кількості спостережень. Недоліки методу пов'язують із залежністю результату від вибору випадкових початкових позицій кластерних центрів, чутливістю до викидів, а також необхідністю введення кількості кластерів, що повинна бути заздалегідь визначена дослідником.

EM-алгоритм заснований на методиці ітеративного обчислення оцінок максимальної правдоподібності. В основі ідеї EM-алгоритму лежить припущення, що досліджувана множина даних може бути змодельована за допомогою лінійної комбінації багатовимірних нормальних розподілів, а метою є оцінка параметрів розподілу, які максимізують логарифмічну функцію правдоподібності, використовуювану як міру якості моделі. З урахуванням цього припущення можна визначити параметри – математичне сподівання і дисперсію, які відповідають закону розподілу елементів у кластері, найкращим чином прийнятному для спостережуваних даних. В результаті маємо, що будь-яке спостереження належить до всіх кластерів, але з різною ймовірністю. Серед переваг EM-алгоритму можна виділити: потужну статистичну основу; стійкість до шумів і пропусків в даних; можливість побудови бажаного числа клас-

терів; швидку збіжність при вдалій ініціалізації. До недоліків відносять такі: припущення про нормальність всіх вимірювань даних не завжди виконується; при невдалій ініціалізації збіжність алгоритму може виявитися повільною; алгоритм може зупинитися в локальному мінімумі.

Практичні результати кластеризації для банків Харківського регіону

Для дослідження вибрано сукупність банків Харківського регіону та їх показники діяльності станом на 01.01.2013 [8]. На цей час у Харківському регіоні діють 25 банків, в якості показників кластеризації використані значення компонентів вектора K виду (4). Вхідні дані для 10 банків наведено у табл. 1.

Далі виконана нормалізація компонентів K , після чого застосовано пакет Statistica 10.0 для кластеризації даних із заданим числом кластерів 2. Оптимальне число кластерів для довільної сукупності

даних можна визначити з використанням спеціальних засобів пакету Statistica [7].

Для нормалізації даних використано Z -показник [2, 5] (стандартне відхилення від середнього). При використанні цього способу обчислюється, на скільки середньоквадратичних відхилень змінна в ряду даних вища або нижча середнього. Це приводить значення показників до стандартного нормального розподілу (середнє 0, стандартне відхилення 1). Додатний Z -показник вказує на те, що значення є вищим за середнє, а від'ємний – що значення є нижчим за середнє.

Для визначення відстані між елементами-векторами показників використана евклідова метрика. Частина вхідних нормалізованих даних та результати кластеризації у вигляді номера кластера та відстані від центра кластера (показник якості кластеризації) для методу k -means наведено у табл. 2. Чим більший показник якості, тим краще виконана кластеризація.

Таблиця 1

Показники діяльності банків

	Назва банку	Загальні активи (млн. грн.)	Власний капітал (млн. грн.)	Статутний капітал (млн. грн.)	Чистий прибуток/збиток (млн. грн.)
1	АО «Банк РенесансКапітал»	723,05	234,29	201,40	7,95
2	ПАО «Креди Агриколь Банк»	13175,7	1670,61	1059,67	408,50
3	ВТБ Банк	34212,3	4050,86	5415,78	951,40
4	Банк «Фінанси та Кредит»	22548,4	1883,47	2100,00	3,17
5	Райффайзен Банк Аваль ПАО	47694,5	6346,03	3002,57	33,85
6	Альфа-Банк	25588,8	4102,07	4653,21	36,25
7	ПриватБанк	172428,7	18300,76	14897,56	1532,76
8	Надра Банк	27684,4	3890,36	4029,93	1,01
9	Правекс-банк	5180,86	1262,71	941,92	-302,29
10	Меркурій	1708,52	221,15	141,50	1,02

Таблиця 2

Компоненти вектора K та результат кластеризації

	Кластер	Назва банку	МК	ІСВК	ПК	Відстань до центра
1	1	АО «Банк Ренесанс-Капітал»	-1,6068	-0,0264	-0,0995	0,2356
2	2	ПАО «Креди Агриколь Банк»	0,6173	-0,5457	1,9858	0,3537
3	2	ВТБ Банк	0,8762	1,0737	1,8902	1,0486
4	2	Банк «Фінанси та Кредит»	2,5098	0,5621	-0,4189	1,0893
5	2	Райффайзен Банк Аваль ПАО	0,4454	-0,9172	-0,3827	1,0422
6	1	Альфа-Банк	-0,1465	0,6067	-0,3481	1,0213
7	2	ПриватБанк	1,3285	-0,1315	0,3938	1,0108
8	2	Надра Банк	0,2603	0,3798	-0,4330	1,0223
9	1	Правекс-банк	-1,1357	-0,2884	-2,8062	1,1146
10	2	Меркурій	0,5426	-0,5330	-0,3900	1,0210

Як бачимо із наведеного фрагменту даних табл. 2, банки АО «Банк Ренесанс-Капітал», Альфа-Банк, Правекс-банк віднесено до кластеру 1, решта 7 банків – до кластеру 2. Загалом, за результатами кластеризації методом k -means 11 банків були віднесені до кластера 2, який відзначається дещо кращим рівнем капіталізації. Висновок про вищий рівень капіталізації групи 2 з аналізованих 25-ти банків можна зробити також безпосередньо із значень табл. 1, 2 та на підставі даних [8].

Додатковий роздільний аналіз показників МК, ІСВК, ПК для банків регіону показує, що показник

МК має найбільше різноманіття та діапазон значень і визначає, головним чином, кінцеве розбиття банків на групи. Значення решти показників (ІСВК, ПК) достатньо близькі між собою для всіх 25 банків, тому вони практично не впливають на кінцевий результат. Діаграма значень показника МК для усіх 25 банків регіону наведена на рис. 1.

У результаті розбиття числа банків на два кластери 1–2 склало: 14–11 (метод k -means) та 19–6 (метод EM) відповідно у першому та другому кластерах. Загалом, результати кластеризації двома методами для конкретних банків співпали для 13 об'єктів

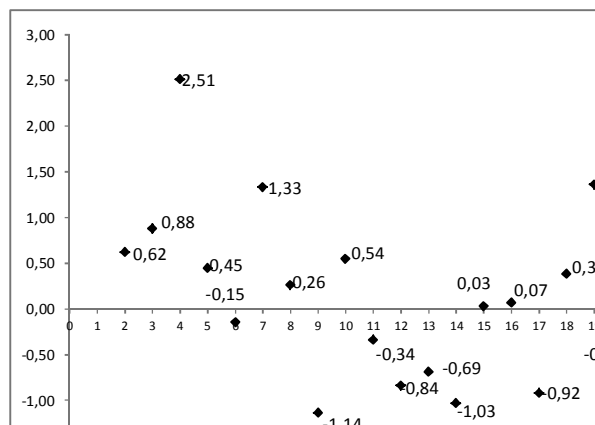


Рис. 1 Значення показника МК

із 25, що підтверджує об'єктивність проведеного оцінювання. Порівняння результатів кластеризації методами k-means та EM показує, що при використанні підходу EM 5 банків із кластеру 2 перейшло до кластеру 1, тобто, перейшло до групи менш ефективних за обговорюваними показниками. Цей результат можна вважати зрозумілим, так як метод EM виявився більш чутливим до неперервного змінювання аналізованих даних. Результати групування залежать від використаного методу кластерного аналізу, так як кожен метод базується на своєму математичному апараті.

Висновки

Впровадження методів аналізу даних при підтримці прийняття ефективних економічних рішень сприяє виявленню скритих закономірностей з метою покращення кінцевих результатів діяльності. Результати кластеризації банків на групи в цілому залежать від використаного методу кластерного аналізу, зважаючи на різний апарат моделей оброблення.

Наукова новизна дослідження полягає у застосуванні інтелектуальних інформаційних технологій аналізу даних для аналітичного оброблення банківської економічної інформації. Практична значущість роботи – на прикладі показників діяльності банків

Харківського регіону показана можливість виявлення реальних закономірностей у результатах економічної діяльності, що в цілому сприяє підвищенню ефективності прийняття менеджерських рішень.

Перспектива досліджень полягає в розширенні системи показників з метою забезпечення універсальності та узагальненості прийняття рішення.

Список літератури

1. Гусева Ю.Ю. Диагностика рівня капіталізації банківської установи / Ю.Ю. Гусева, С.В. Байталюк // *Економіка та управління підприємствами машинобудівної галузі: проблеми теорії та практики*. – 2012. – № 2(18). – С. 75-82.
2. Васюренко О.В. Економічний аналіз діяльності комерційних банків: навч. пос. / О.В. Васюренко, К.О. Волохата. – Х.: ХФ УАБС, 2003. – 396 с.
3. Гороховатський В.О. Розроблення програмного середовища для оперативного розрахунку фінансово-кредитних показників / В.О. Гороховатський, А.М. Кобылін, Р.В. Семенов // *Вісник Університету банківської справи Нац. банку України*. – 2013. – №1 (16). – С. 270-273.
4. Barylyuk I. Cluster analysis of capitalization of Ukrainian banks in the conditions of national economy globalization / I. Barylyuk, N. Paitra, M. Yastrubskyy // *Econtechmod. An international quarterly journal*. – 2013. – Vol. 2. No. 1. – P. 3-14.
5. Прикладная статистика: классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.
6. Методы и модели анализа данных: OLAP и Data Mining: учебн. пос. / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб: БХВ-Петербург, 2004. – 336 с.
7. Statistica [Електронний ресурс] / Офіційний сайт компанії Statsoft. – Режим доступу до ресурсу: <http://www.statsoft.ru/products>. Загол. з екрану.
8. Структура активів банків України. Прибутки і збитки банків України. Структура власного капіталу банків України за станом на 01.01.2013 (у розрізі банків) // *Вісник Нац. банку України*. – 2013. – №3 (205). – С. 50-70.

Надійшла до редколегії 16.12.2013

Рецензент: д-р фіз.-мат. наук О.В. Ситнік, Інститут радіофізики і електроніки НАН України, Харків.

ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ АНАЛИЗА ДАННЫХ ДЛЯ ОБРАБОТКИ ИНФОРМАЦИИ О ДЕЯТЕЛЬНОСТИ БАНКОВ

В.А. Гороховатский, А.М. Кобылин, Р.В. Семенов

Обсуждается проблема оценки эффективности работы банков на основе анализа значений системы показателей. Применение методов кластерного анализа векторных данных уровня капитализации позволяет выделить группу банков с целью выявления общих тенденций их развития и дальнейшего управления процессом их функционирования. Приведены результаты анализа данных деятельности банков Харьковского региона.

Ключевые слова: анализ деятельности банков, система показателей, кластеризация, показатели капитализации, нормализация компонентов, принятия решений.

APPLICATION OF INTELLIGENT TECHNOLOGIES FOR DATA PROCESSING OF INFORMATION ABOUT THE ACTIVITIES OF BANKS

V.A. Gorohovatsky, A.M. Kobilyn, R.V. Sementsov

The problem of evaluating the effectiveness of the bank based on the analysis of performance. Application of cluster analysis of vector data on capitalization makes it possible to identify a group of banks to identify general trends in the development and subsequent management of their operation. The results of the analysis of data on the activities of banks in the Kharkiv region.

Keywords: analysis of banks, scorecard, clustering, capital ratios, the normalization of the components, decision-making.