

УДК 519.2

Ю. І. Мінаєва

Київський національний університет будівництва і архітектури, Київ

ВИБІР ПОРТФЕЛЯ ЦІННИХ ПАПЕРІВ НА ОСНОВІ МОДЕЛІ ЙОГО ВІДПОВІДНОСТІ РИНКУ

Розглянуто задачу вибору портфеля цінних паперів, подібного за поведженням до поведження ринку цінних паперів. Розв'язок поставленої задачі запропоновано виконувати шляхом побудови та порівняння бінарних матриць, що характеризують структуру ринку і обраного портфеля. Наведено приклад.

Ключові слова: портфель цінних паперів, вибір портфеля цінних паперів, кластеризація, дендрограма, бінарне дерево

Вступ

Один із шляхів дослідження ринку цінних паперів (ЦП) як складової економічної науки в сьгодні полягає в тому, що, перш за все, потрібно враховувати хаотичну природу ринку, що відповідає його поведженню і зумовлює необхідність застосування нових (зокрема, фрактальних) принципів для моделювання його поведження. Показано [1], що для ефективної роботи на ринку ЦП потрібно враховувати основну структуру ринку і тільки тоді, коли можемо розкрити *основну структуру ринку* й привести у відповідність наше бачення процесів, що відбуваються в дійсності, одержимо результат – успіх на ринку.

Згідно цієї парадигми, задача вибору портфеля цінних паперів розглядається як визначення такої підмножини ЦП, структурні властивості якої збігаються з властивостями вихідної множини ЦП (ринку). Ця задача може бути сформульована в термінах кластерного аналізу як побудова *подібних бінарних дерев*. Це означає застосування нової стратегії формування портфеля ЦП – забезпечити поведження портфеля, аналогічне (з погляду зльотів і падінь) поведженню ринку. На жаль, питання подоби ієрархічних структур розроблені досить приблизно, загальна теорія практично відсутня, тому в більшості випадків застосовують моделювання, пошук непрямих аналогій, безпосередню перевірку отриманих теоретичних результатів на практиці.

Щодо поставленої задачі питання в термінах кластерного аналізу стоїть так: чи можна (і як) проводити кластеризацію в просторі меншої вимірності, ніж вихідний, не втрачаючи при цьому певних властивостей вихідного простору? Для відповіді треба зробити деякі уточнення.

Скорочення простору вигідно, насамперед, за семантичними причинами: в процесі скорочення простору обираються найбільш важливі інформативні характеристики, що істотно саме по собі [9;11]. Результати кластеризації в скороченому просторі стійкіші і надійніші, ніж у вихідному багатовимірному; мала кількість параметрів легше піддається змістовному сприйняттю і подальшому аналізу, ніж велика; у випадку скорочення простору до *дво- або тривимірного* дані стають візуально спостережуваними, а наочність корисна завжди. Крім того, скорочення числа ознак (особливо при розв'язанні задач надвисокої вимірності) призводить до суттєвого спрощення обчислювальних процедур і підвищує ступінь технічної реалізованості проблеми.

Однак, при цьому виникають нові задачі, що не поступаються складністю вихідній: потрібно визначити, що саме необхідно перенести з вихідного простору неспотвореним (або малоспотвореним) у скорочений. Показано, що при цьому бажано зберегти *структуру* (дуже часто геометричну) множини об'єктів, для цього існують різні критерії – об'єктивні й суб'єктивні [1, 2-6].

Постановка задачі

Постановка задачі вибору портфеля на підставі кластерного аналізу може бути такою: є дві множини ЦП – $S^{(1)}$ і $S^{(2)}$, $S^{(1)} \supseteq S^{(2)}$, що володіють однаковою системою ознак: $p_1, p_2, \dots, p_k, \dots$. Для множин $S^{(1)}$ і $S^{(2)}$ побудовані дендрограми – $D^{(1)}$, $D^{(2)}$ відповідно, причому кількість термінальних вузлів $\{n_i^{(i)}\}$, $i=1, 2, \dots, N^{(i)}$ в $D^{(1)}$, $D^{(2)}$ однакова, для дендрограм $D^{(1)}$, $D^{(2)}$ побудовані 2-адичні матриці $M^{(1)}$, $M^{(2)}$ з вимірностями $r^{M1 \times N^{(1)}}$, $r^{M2 \times N^{(2)}}$.

Умова $|\Gamma^{M1} - \Gamma^{M2}| \leq \delta_r$ є необхідною умовою близькості.

Достатні умови близькості множин $S^{(1)}$ і $S^{(2)}$ можна сформулювати у такому вигляді:

$$1) \left| \left[M^{(1)} \right]_{N^{(1)}}^{\Gamma^{M1}} - \left[M^{(2)} \right]_{N^{(2)}}^{\Gamma^{M2}} \right| \rightarrow \min;$$

$$2) |\aleph_1([M^{(1)}]_{N^{(1)}}^{\Gamma^{M1}}) - \aleph_2([M^{(2)}]_{N^{(2)}}^{\Gamma^{M2}})| \leq \delta_r,$$

де $\aleph_1(M^{(1)}), \aleph_2(M^{(2)})$ – 2-адичні числа, обчислені на підставі матриць $M^{(1)}, M^{(2)}$; зазначимо, що обидві умови еквівалентні. Величина δ_r – значення умови на близькість дендрограм.

Для 2-адичних матриць $M^{(1)}, M^{(2)}$ з вимірностями $\Gamma^{M1} \times N^{(1)}, \Gamma^{M2} \times N^{(2)}$ можуть бути визначені мінімальні шляхи - $\wp^{(1)}$ і $\wp^{(2)}$. Якщо ці шляхи близькі (в будь-якій метриці), тобто $\wp^{(1)} \approx \wp^{(2)}$, то ця умова може бути додатковою достатньою умовою, яку можна використовувати для наближених оцінок.

Дослідження показують, що гарантувати однаковість поведінки ринку ЦП і портфеля у всіх умовах на підставі сформульованих оцінок (необхідності і достатності) немає повних підстав, хоча практичні обчислення показують, що для більшості розглянутих випадків адекватність стратегії гарантується. Однак можна запропонувати додаткові перевірки, які засновані на застосуванні двоїстості в задачах класифікації.

Припустимо, що ринок складається з 1000 ЦП, для яких відомі вартість, прибутковість, ризик, портфель складається з 100 ЦП із такими ж параметрами. Відповідно до постановленої задачі ринок кластеризується, наприклад, на 50 кластерів (по 20 ЦП), портфель повинен мати також 50 кластерів (по 2 ЦП), тому що тільки в цьому випадку гарантується однаковість (або близькість) ієрархічних структур. Після кодування отриманих дерев обчислюються значення матриці C і вектора x для віртуального ринку і портфеля – C_{R^x} і C_{P^x} відповідно. Якщо $|C_{R^x} - C_{P^x}| \leq \delta_x$, то структурно віртуальний ринок і портфель близькі, що спричиняє їх адекватне поведіння. Можна використовувати менш строгу оцінку - $|\Gamma_R^{\max} - \Gamma_P^{\max}| \leq \delta_r$, де $\Gamma_R^{\max}, \Gamma_P^{\max}$ - максимальні ранги (відстані) бінарних дерев, що відповідають ринку і портфеля.

Подібність бінарних дерев. Відомо, що кластеризація множини об'єктів $X = \{X_i\}, i=1, n$ може бути виконана таким чином, що кожний кластер може містити 1 об'єкт, 2 об'єкти, ..., n-1 об'єктів. Випадок, коли множина об'єктів $X = \{X_i\}, i=1, n$; кластеризована у вигляді «один об'єкт – один

кластер», вважається такою, коли ієрархічні властивості об'єкта проявляються найбільш повно. Підставою для такого твердження є те, що з усіх кластеризацій величини C_x і максимальні ранги бінарних дерев максимальні. Сформульовано твердження.

Твердження. Нехай виконані кластеризації об'єкта $X = \{X_i\}, i=1, n$ по одному з m елементів, X_1 і $X_i, i=1, m, m < n$, і побудовані бінарні дерева $\aleph^{(1)}$ і $\aleph^{(m)}$, для яких визначені максимальні ранги - $\Gamma^{(\aleph_1)}$ і $\Gamma^{(\aleph_m)}$ відповідно. Будемо вважати кластеризації адекватними (близькими), якщо $\Gamma^{(\aleph_1)} \approx \Gamma^{(\aleph_m)}$, відношення $\Gamma^{(\aleph_1)} / \Gamma^{(\aleph_m)}$ характеризує ступінь адекватності.

Зазначимо, що питання адекватності дендрограм розглянуті в [7;9], де введений коефіцієнт адекватності, що є евристичним і істотно наближеним. Запропонований у роботі спосіб є більш об'єктивним, тому що використовує формально обчислену величину – C_x або максимальний ранг дендрограми. Однак теоретичні основи визначення адекватності дендрограм, отриманих на основі кластеризації, дотепер практично не розглянуті і основою для ухвалення рішення є експеримент. На рис.1 наведено об'єкт $X = \{X_i\}, i=1, n$, що містить 100 елементів. На рис. 2 показані результати його кластеризації за методом найбільш віддаленого сусіда (рис.2,а) і за методом найближчого сусіда (рис. 2,б). На осі X вказано номер об'єкта, на осі Y вказано відповідну характеристику об'єкта, для даного випадку це прибутковість ЦП.

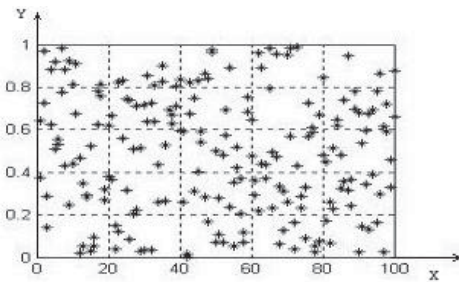


Рис. 1. Вихідна множина об'єктів

Розглянемо отримані дендрограми: по-перше, варто визнати, що розглядаючи послідовність дендрограм: $25 \times 100, 10 \times 100 (100 \rightarrow 50 \rightarrow 25 \rightarrow 10)$, можна бачити, що у всіх випадках вони можуть аналізуватися на рівні використання системи стискаючих відображень⁹ з усіма властивостями, що впливають з цієї обставини. Для побудованих на

⁹ Запис 25×100 означає отримання 25-ти кластерів з 100 елементів

рис. 2 дендрограм на осі X позначають номер об'єкта, а на осі Y – ранги дендрограмми.

Як видно з рис. 2,а ці стискання практично в жодному випадку не спотворюють структуру вихідної множини до рівня кластеризації по 10 елементів у кластері, максимальний ранг дерева змінюється в межах 10 %. У випадку, наведеному на рис. 2,б вже в другому стисканні (2 елементи в кластері) при збереженні зовнішньої подоби структурні властивості об'єкта істотно (на 30%) зменшуються, при кластеризації по 10 елементів властивості зменшуються в 2 рази. Однак хоча структурні властивості об'єкта зменшуються при збільшенні кількості елементів в кластері, інтегрально вони не зникають. Геометричну аналогію можна навести таку: циліндр об'ємом в 100 л при його деформації (наприклад, вертикальному стиску) може мати об'єм в 50, 25, 10л і т.д., зберігаючи при цьому властивості об'єкта (обчислення об'єму за однією й тією ж формулою).

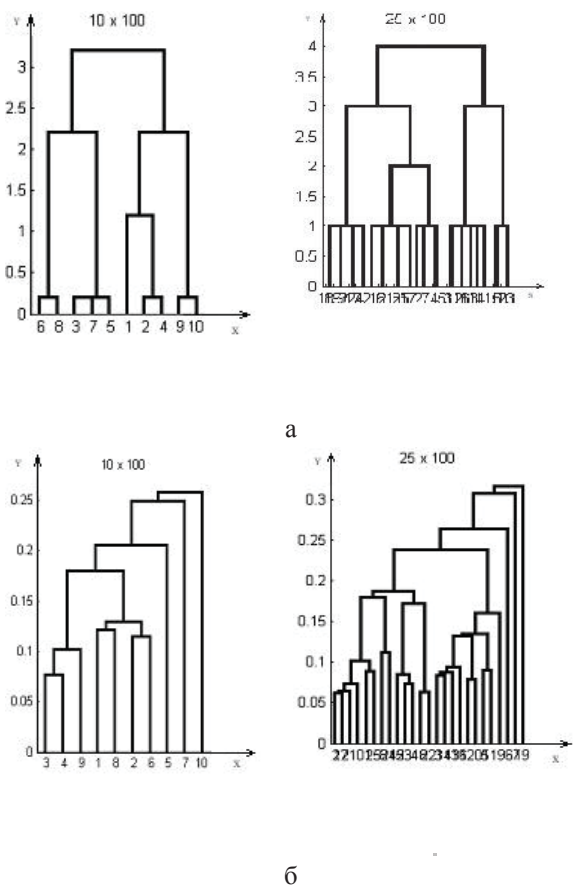


Рис.2. Дендрограми кластеризації по об'єктах $X = \{x_i\}$, $i=1, 100$: а) – метод найбільш віддаленого сусіда; б) – метод найближчого сусіда.

На рис. 3,а і 3,б залежності максимальних рангів дерева від кількості елементів у кластері: а - метод найбільш віддаленого сусіда: $y = 0.032*x^3 - 0.28*x^2 + 0.92*x + 2.8$ –

б- метод найближчого сусіда: $y = 0.032*x^3 - 0.22*x^2 + 0.51*x - 0.08$. Отримані залежності підтверджують тезу про афінну подобу отриманих дерев.

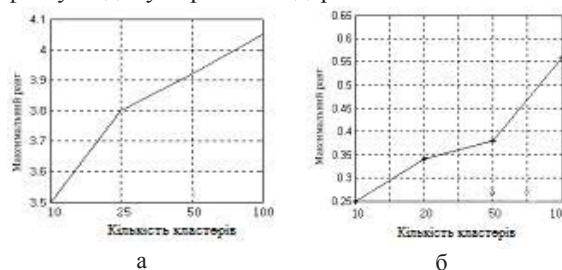


Рис.3. Залежності максимального рангу від кількості елементів у кластері: а - метод найбільш віддаленого сусіда, $y = 0.032*x^3 - 0.28*x^2 + 0.92*x + 2.8$; б - метод найближчого сусіда, $y = 0.032*x^3 - 0.22*x^2 + 0.51*x - 0.08$

На рис. 4 наведена множина ЦП, що імітує ринок в координатах «вартість-прибутковість-ризик», $X = \{x_{ijk}\}$, $i, j, k=1, 50$ (значення координат нормалізовані, вартість – 10^6 у.о., прибутковість - %, ризик - % і його дендрограми для різних типів кластерів).

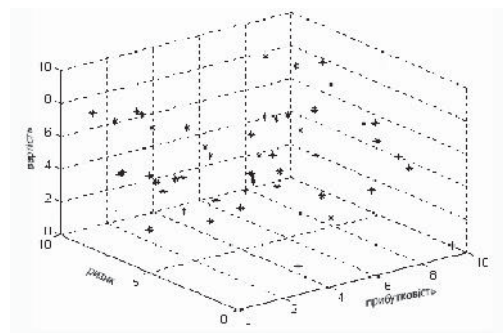
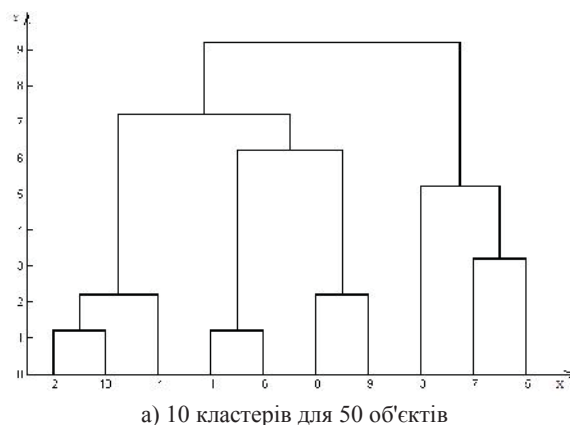


Рис. 4. Вихідна множина точок у тривимірній системі координат, що імітує ринок ЦП

На рис. 5 наведені дендрограми множини ЦП, отримані за методом найбільш віддаленого сусіда.



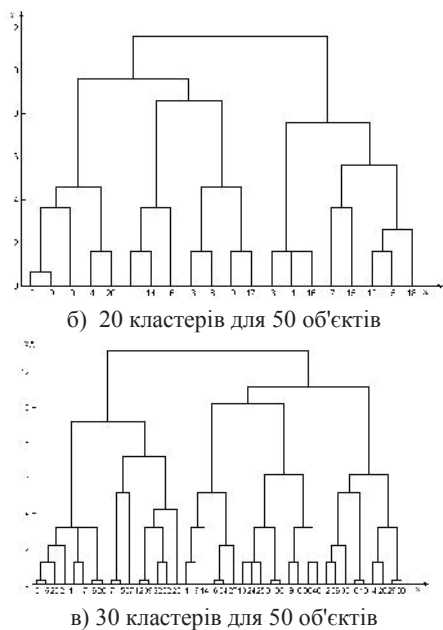


Рис.5. Дендрограми множини ЦП, отримані за методом найбільш віддаленого сусіда

Розглянемо дендрограми множини ЦП, отримані за методом найбільш наближеного сусіда (рис. 6).

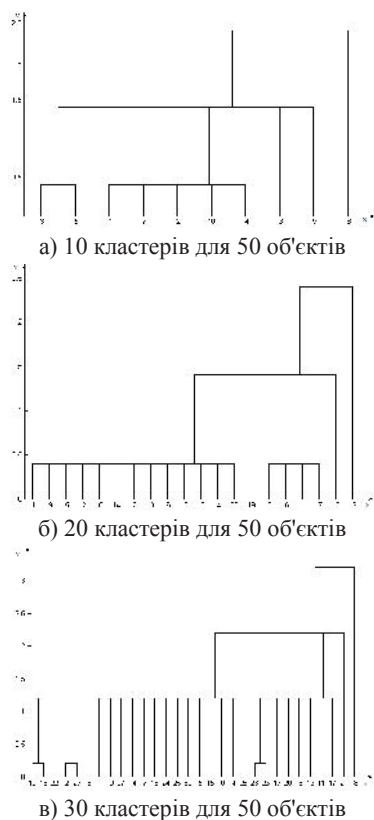


Рис. 6. Дендрограми множини ЦП, отримані за методом найбільш наближеного сусіда

Умовне подання ієрархічної кластеризації як процедури послідовної реалізації стискаючих відображень 30×50 , 20×50 , 10×50 показане на рис.7.

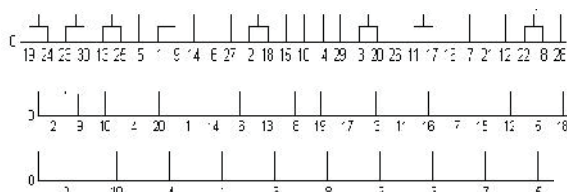


Рис. 7 Подання ієрархічної кластеризації як процедури послідовної реалізації стискаючих відображень: 30×50 , 20×50 , 10×50

Аналізуючи наведені зображення дендрограм, отриманих за методами найбільш віддаленого й найбільш наближеного сусідів, можна бачити афінну подоби всіх дендрограм, що практично не залежить від ступеня вкладеності об'єктів у кластер. Користувач може застосовувати практично будь-яку кластеризацію, пам'ятаючи, що властивості початкової дендрограми (один елемент - один кластер) проявляються в *стислій* дендрограмі прямо пропорційно співвідношенню $\Gamma_{l \times m}^{\max} / \Gamma_{n \times m}^{\max}$, де $\Gamma_{l \times m}^{\max}$ - максимальний ранг дендрограми типу „1 кластер – один елемент”; $\Gamma_{n \times m}^{\max}$ - максимальний ранг дендрограми типу «1 кластер – n елементів»; m - загальна кількість елементів в об'єкті.

Питання ефективного кодування дендрограм є досить складними і в теоретичному плані дотепер практично не вирішені [8;9]. В роботі автором прийнята ідеологія незалежності результату від виду й форми кодування. Це досягнуто тим, що ринок ЦП і портфель ЦП, представлені у вигляді дендрограм, закодовані в однаковій формі (можливо, навіть недостатньо ефективній з погляду на деякі критерії).

Як відомо, питання чи представляє дендрограма основну ієрархічну структуру даних дотепер залишається відкритим. В *p*-адичному базисі, проте, немає неоднозначності щодо інтерпретації дендрограм. Причина в тому, що *p*-адична відстань – ультраметрика і це визначає те, що *p*-адична дендрограма *правильно представляє ієрархію* в межах даної множини *p*-адичних чисел відповідно до метрики[17;18]. Інший ефект полягає в тому, що *p*-адична класифікація алгоритмічно значно більш проста, ніж класичний аналог. Послідовність для вилучення даних лежить в операціях зрушення для класифікації закодованих даних.

Якщо дендрограма X відома, тоді *p*-адичне кодування може бути ефективним через асоціацію шляху від верхніх кластерів до даних з *p*-адичними числами. Матрична форма *p*-адичного кодування така: $x = Cr$, де x - десяткове кодування, C - матриця з кодами дендрограмних переходів; r - вектор ступенів фіксованого цілого (зазвичай, більш обмежно, фіксованого простого p).

Експериментальне дослідження ефективності запропонованої методології

Експериментальне дослідження ефективності запропонованої методології аналізу відповідності портфеля ЦП ринку і вибору раціонального портфеля включає такі задачі [10-12]:

- оцінка зміни структури множини під час коливань окремих параметрів;
- оцінка відповідності портфельної множини ЦП ринку і модифікація портфеля на підставі оцінок фрактальних вимірностей, максимальних рангів і величини адичної матриці.

Три тестових множини ЦП, що складаються кожна зі 100 елементів, у координатах «ризик-прибутковість», показані на рис.8. Прийняті позначення: * – вихідна множина; +, Δ – коливання параметрів у межах ± 5%.

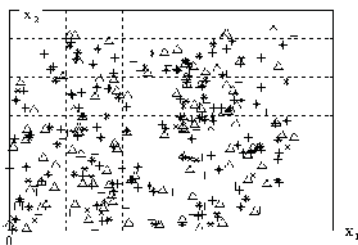
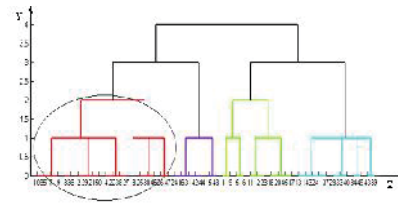
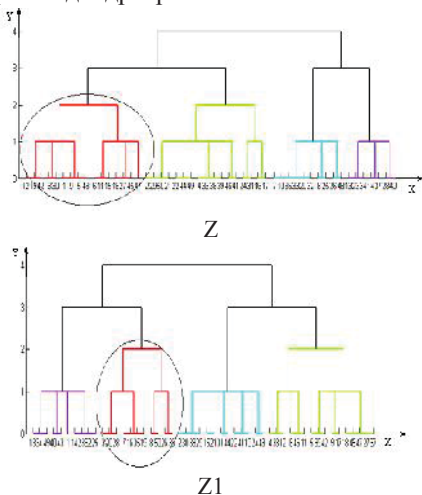


Рис.8. Характеристики ЦП у координатах $\{x_1, x_2\}$

На рис.9 наведені дендрограми зазначених множин: Z – дендрограма вихідної множини; Z1 і Z2 – дендрограми множин з коливаннями параметрів у межах ± 5%. Всі дендрограми побудовані за методом найбільш віддаленого сусіда (50 кластерів на 100 елементів). Для побудованих на рис. 9-11 дендрограм на осі X позначають номер об'єкта, а на осі Y – ранги дендрограми.



Z2

Рис.9. Дендрограми тестових множин: Z – дендрограма вихідної множини; Z1 і Z2 – дендрограми множин з коливаннями параметрів у межах ± 5%

Аналізуючи дендрограми, можна бачити, що під час кодування алфавітом $\{0;1\}$ вони являють собою матриці вимірністю 50×4 , максимальні ранги у всіх випадках практично однакові, але істотно змінився склад кластерів на всіх рівнях, починаючи з $v=1$. Такий важливий висновок полягає в тому, що коливання параметрів множини ЦП (зміна цін, ступеня ризиків і ін.), що імітує ринок, привели до значної зміни його структури, зберігши зовнішні параметри практично незмінними.

Окремо розглянемо виділені на рис. 9 кластери. В ці кластери потрапили елементи, в яких найбільші відстані, зміни параметрів моделі, привели до того, що кількість елементів в об'єднаному кластері для $v=1$ і $v=2$ істотно змінилася, $n_{Z1} < n_Z < n_{Z2}$, що добре видно з рис.10.

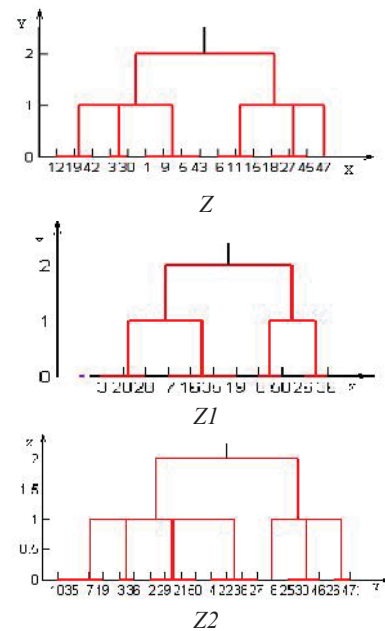


Рис.10. Кластери, в які потрапили елементи з найбільшими відстанями

На рис. 11 наведені дендрограми тестових множин: Z – дендрограма вихідної множини; Z1 і Z2 – дендрограми множин з коливаннями параметрів в межах 5% при кластеризації по 10 елементів у кожному кластері.

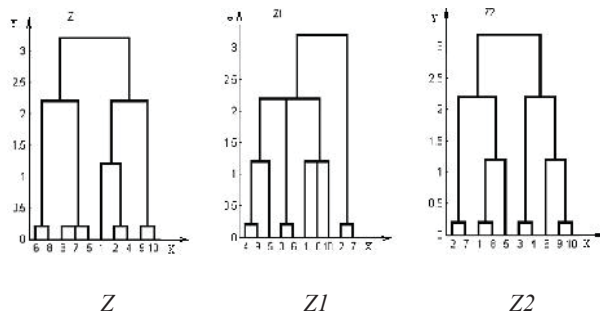


Рис.11. Дендрограми тестових множин: Z – дендрограма вихідної множини; Z1 і Z2 – дендрограми множин з коливаннями параметрів в межах 5%

Автором розроблене програмне забезпечення для визначення розподілу елементів тестових множин по кластерах [10-12]. Аналізуючи дендрограми, можна бачити, що, як і в попередньому випадку, при кодуванні алфавітом {0,1} вони являють собою матриці вимірністю 10×3, максимальні ранги у всіх випадках практично однакові, але істотно змінився склад кластерів на всіх рівнях, починаючи з $v=1$.

Підтверджується висновок про те, що коливання параметрів множини ЦП (зміна цін, ступеня ризиків і ін.), що імітує ринок, призводять тільки до зміни його ієрархічної структури, зберігши зовнішні параметри практично незмінними. Задача визначення відповідності портфеля ринку і за наявності невідповідності визначення рекомендації (щодо складу й структури портфеля) для одержання відповідності розглядається нижче. На рис.12 наведені тестові множини ЦП, що імітують портфель і ринок.

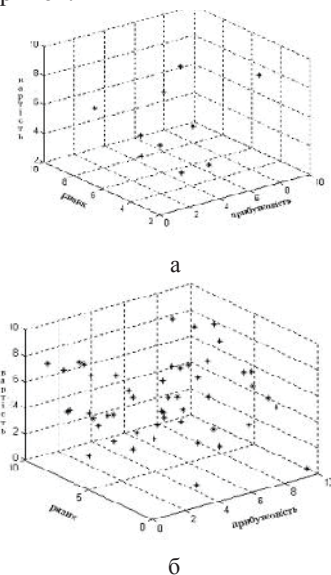


Рис.12 Тестові множини ЦП, що імітують портфель (а) і ринок (б): необхідно визначити відповідність портфеля і ринку, за наявності невідповідності визначити рекомендації (щодо складу і структури портфеля) для одержання відповідності

На рис.13 представлені закодовані дендрограми для множин ЦП, що імітують ринок і портфель.

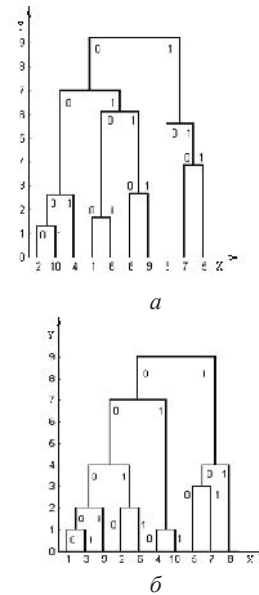


Рис.13. Дендрограми (МНВС) для множин ЦП, що імітують ринок (а) і портфель (б)

На рис.14 наведені бінарні матриці, що відповідають дендрограмам, наведеним вище. У прийнятих матрицях кількість рядків відповідає кількості змінних (елементи об'єкта), кількість стовпців - кількість рівнів (ранги). На підставі аналізу матриць робиться висновок про відповідність підмножин - ринок-портфель. Для такої оцінки використовуються кілька методів:

– обчислення фрактального числа 2 – адично закодованих матриць $C = \sum_{level} c_v 2^v \in Q_2$,

де коефіцієнти $c_v \in [0, 1]$ – числа, визначені на рівні v ;

– обчислення фрактальних вимірностей матриць;

– обчислення сум рангів матриць: сума рангів стовпців – $sum(C)$, сума рангів рядків - $sum(C^T)$, T - символ транспозиції, відносна сума рангів – $sum(sum(C))/max(sum(C))$.

	1	2	3	4	5	6	7	8	9	
1	0	0	0	0	0	0	1	0	0	1
2	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	1
4	0	0	1	0	0	0	0	0	0	0
5	0	0	0	1	0	1	0	0	0	1
6	0	1	0	0	0	0	1	0	0	0
7	0	0	0	0	0	1	0	0	0	1
8	0	0	0	0	0	1	1	0	0	0
9	0	0	1	0	0	1	1	0	0	0
10	1	0	0	0	0	0	0	0	0	0

а

1	2	3	4	5	6	7	8	9	
0	0	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	2
1	0	0	0	0	0	0	0	0	3
0	0	0	0	0	0	1	0	0	4
0	0	0	0	0	0	0	0	1	5
0	1	0	1	0	0	0	0	0	6
0	0	1	0	0	0	0	0	1	7
0	0	0	1	0	0	0	0	0	8
0	1	0	0	0	0	0	0	0	9
1	0	0	0	0	0	0	0	0	10

Рис.14. Матриці дендрограм: а – ринок, б –портфель

Зазначимо, що обчислення сум рангів для р-адичного обчислення відповідають переходу до р-адичних оцінок, де використовують тільки порядки змінних. Обчислення фрактального числа 2-адично закодованих матриць виконується в такий спосіб: а) $x_1=2^7, x_2=0, x_3=2^9, x_4=2^3, x_5=2^4+2^6, x_6=2^2+2^7, x_7=2^6+2^9, x_8=2^6+2^7, x_9=2^3+2^6+2^7, x_{10}=2^1, C_M=\sum_1^{10} x_i$; б) $x_1=0, x_2=2^4, x_3=2^1, x_4=2^7, x_5=2^9, x_6=2^2+2^4, x_7=2^3+2^9, x_8=2^4+2^9, x_9=2^2, x_{10}=2^1, C_P=\sum_1^{10} x_i$.

Результати експериментів наведені в таблиці.

Таблиця

Дослідження близькості підмножин ЦП, що моделюють портфель і ринок

Тип експерименту	Об'єкти		
	Ринок	Портфель	Похибка обчислень, %
Обчислення 2-адичних чисел матриць	1830.0	1732.0	5.36
Фрактальна вимірність	1.20	1.08	10.38
Обчислення відносної суми рангів - $\frac{\text{sum}(\text{sum}(C))}{\text{max}(\text{sum}(C))}$	4.0	4.0	0

Таким чином, представлена підмножина ЦП, розглянута як портфель, має структуру й, природно, поведження, аналогічні зі структурою й поведженням підмножини ЦП, розглянутої як ринок. Зазначимо, що умова перевірки на рівні бінарних матриць застосовна тільки в тому випадку, коли максимальні ранги дендрограм, що моделюють портфель і ринок, близькі (різниця не перевищує 0 %). В іншому випадку необхідно змінювати склад портфеля, намагаючись досягти збігів значень максимальних рангів дендрограм.

Висновки

1. Сформульовано новий клас задач – визначення портфеля ЦП, поведження якого буде схожим на поведження ринку ЦП у цілому, якщо під поведженням розуміти зростання і падіння

вартості ЦП. В свою чергу, сформульований клас задач передбачає:

- визначення того, наскільки існуючий портфель схожий на ринок;

- визначення способу, за допомогою якого можна зробити портфель схожим на ринок.

2. Задача вибору інвестиційного портфеля, розглянутого як підмножина ЦП, властивості якого збігаються із властивостями вихідної множини ЦП (ринок), в термінах кластерного аналізу сформульована як побудова близьких (подібних) бінарних дерев.

Список літератури

1. Вильямс.Б. Торговый хаос. Экспертные методы максимизации прибыли.- М.: Аналитика, 2000. – 198 с.
- 2.Дюран Н., Оделл П. Кластерный анализ.-М.: Статистика, 1977.-128 с.
- 3.Диде Э. и др. Методы анализа данных.-М.: Финансы и статистика, 1985.- 360 с.
- 4.Мандель И. Д. Кластерный анализ.- М.: Финансы и статистика. 1988.-176 с.
- 5.Simon H.A.. The Sciences of the Artificial. MIT Press, Cambridge, MA, 1996.
- 6.Бериков В.Б., Лбов Г.С. Современные тенденции в кластерном анализе.– Ин-т матем. им.С.Л.Соболева СО РАН.–Интернет-ресурс:http://window.edu.ru/window_catalog/pdf2txt?p_id=27124
- 7.Bradley P.E. Mumford dendrograms. Computer Journal, 2007. – p.1-12.
- 8.Murtagh F., Downs G., and Contreras P.. Hierarchical clustering of massive, high dimensional data sets by exploiting ultrametric embedding. SIAM Journal on Scientific Computing. Vol. 30, No. 2, pp. 707–730. March 2008.
- 9.Murtagh F. Identifying the ultrametricity of time series. Eur. Phys. J. B. 2005. V. 43 , P. 43, 573–579 (2005)
10. Мінаєва Ю.І. Методи нечіткої ієрархічної кластеризації в задачах формування інвестиційного портфелю акцій: Матеріали міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми вычислительного интеллекта (ISDMCI'2010)»
11. Мінаєва Ю.І. Визначення портфеля цінних паперів на основі інваріантних підмножин ієрархічної кластеризації. Системний аналіз та інформаційні технології: Матеріали XII Міжнародної науково-технічної конференції (25-29 травня 2010 р., Київ). – К.: ННКА «ІПСА» НТУУ «КПІ», 2010. – 290 с.
12. Мінаєва Ю. І. Кластерний аналіз в задачах вибору портфеля цінних паперів. Наукова конференція молодих вчених, аспірантів і студентів КНУБА (16-18 листопада 2009р., Київ) : тези доповідей. – в 2-х частинах. – Ч.1. - К.: КНУБА, 2010. – 157 с.

Стаття надійшла до редколегії 10.05.2011

Рецензент: д-р техн. наук, проф. М.К.Печурін, Національний авіаційний університет, Київ