

КЛАСТЕРИЗАЦИЯ ЦЕННЫХ БУМАГ С ПОМОЩЬЮ α - КВАЗИЭКВИВАЛЕНТНОСТИ.

В статті розглядається застосування нечіткого кластерного аналізу для розбиття набору цінних паперів на певні групи, що мають однакові тенденції. Ключові слова: нечіткий кластерний аналіз, α - квазіеквівалентність, кластери, цінні папери (ЦП), портфель цінних паперів.

В статье рассмотрено применение нечёткого кластерного анализа для разбиения набора ценных бумаг на определённые группы, имеющие одинаковые тенденции. Ключевые слова: нечёткий кластерный анализ, α - квазіеквівалентність, кластеры, ценные бумаги (ЦБ), портфель ценных бумаг.

This article is devoted to the unclear cluster analysis wich considered for breaking up of set of securities on certain groups with having identical tendencies. Keywords: unclear cluster analysis, α - kvaziequivalence, clusters, securities , brief-case of securities.

Постановка проблеми. Общеизвестно, что изменение курсовой стоимости и дивидендов различных ценных бумаг не только в Украине, но и во всем мире зависит от ряда микро- и макрофакторов, экономического и неэкономического характера. Эти факторы могут быть взаимосвязаны в различной степени, а тенденции изменения их динамики способны отличаться друг от друга в достаточно сильной степени. Следовательно, изменение стоимости инвестиционного портфеля в результате сложения различных тенденций с большой вероятностью оказывается достаточно сложной и практически непредсказуемой, если использовать обычный регрессионный анализ. Основные факторы влияния сказываются на различные ценные бумаги не только с разной эффективностью, но зачастую и в прямо противоположных направлениях. Например, повышение цен на нефть может благоприятно сказаться на ценных бумагах нефтяных корпораций, негативно отразившись на автомобилестроительном секторе. В свете вышесказанного, перед инвесторами возникают следующие проблемы:

- 1) Определение с максимальной степенью точности существенных факторов и их влияние на курс ценных бумаг;
- 2) Составление научно-обоснованного прогноза динамики поведения этих ценных бумаг, основываясь на изучении данных факторов;

3) Составление на основе полученных сведений о фондовом рынке оптимального инвестиционного портфеля, что позволяет максимизировать прибыль от вложений при заданной степени риска.

Анализ последних достижений и публикаций. Севриновский В.Д., в своих работах исследовал применение чёткого кластерного анализа в задачах инвестиционного программирования, где инвестиционный процесс рассматривался как статистический. Применение нечетких методов кластерного анализа в социально – экономических задачах было исследовано в работах Зайченко Ю.П.

Решённая ранее часть общей проблемы. Как теоретики, так и практики, занимающиеся оптимизацией портфеля ценных бумаг, регулярно сталкиваются с трудностями, когда перед ними возникает практически неизбежно задача разбиения множества существующих ценных бумаг на различные группы с относительно однородной структурой. Проблемным стоит вопрос подбора и согласования выбранных факторов так, чтобы их представление в многомерной системе координат достаточно точно делала разбивку на кластеры, характеризующиеся максимально схожими тенденциями. При этом нужно учитывать, что даже если бы и удалось подобрать точные коэффициенты для существующих количественных факторов, всегда найдутся не менее важные качественные показатели, выразить которые в количественной форме практически невозможно. В связи с этим принято группирование ценных бумаг на основе существующих индустриальных и других классификаций, а также отталкиваясь от априорной доходности (*ex ante*).[2]

Разбиение множества ценных бумаг на отдельные кластеры в зависимости от динамики доходности осуществляется следующим образом: данные по доходности ценных бумаг на протяжении базы прогноза компонуется в общую матрицу вида:

$$\begin{pmatrix} E_{10} & E_{11} & \dots & E_{1T} \\ E_{20} & E_{21} & \dots & E_{2T} \\ \dots & \dots & E_{km} & \dots \\ E_{N0} & E_{N1} & \dots & E_{NT} \end{pmatrix}, \quad (1)$$

где $E_{km} = [E_{km_1}; E_{km_2}]$ доходность по k -ой ЦБ за m -й период, представленный парой чисел: E_{km_1} - нижней границей доходности ЦБ и E_{km_2} - верхней границей доходности ЦБ.

$$k \in [1; N]; \quad m \in [0; T].$$

Далее, разбиение на кластеры происходит через вычисление евклидова расстояния между ценными бумагами p и q по формуле

$$d(E_p, E_q) = \sqrt{\sum_{m=0}^T (E_{pm_1} - E_{qm_1})^2 + \sum_{m=0}^T (E_{pm_2} - E_{qm_2})^2}, \quad (2)$$

де m – номер периода

Изложение основного материала.

Алгоритм кластеризации ценных бумаг с помощью α - квазиэквивалентности

Перед описанием алгоритма необходимо договориться об используемых в нем T - норме и T - конорме. Будем использовать MIN-норму и MAX - конорму.

1. Построить для каждого образца данных нормальную меру сходства по формуле:

$$\mu_{x_p}(x_q) = 1 - \frac{d(x_p, x_q)}{\max_{q \in [1, N]} (d(x_p, x_q))}, \quad p, q = \overline{1, N}, \quad (3)$$

d - расстояние по Евклиду, считается по формуле (2).

2. Построить относительно каждого образца данных на основании нормальной меры сходства относительную меру сходства для пар образцов данных:

$$\xi_{x_p}(x_i, x_j) = 1 - \left| \mu_{x_p}(x_i) - \mu_{x_p}(x_j) \right|, \quad i, j, p = \overline{1, N} \quad (4)$$

3. Построить меру сходства образцов данных на множестве X по формуле:

$$\xi(a, b) = T \left(\xi_{x_1}(a, b), \dots, \xi_{x_N}(a, b) \right) = \min_{i=1, N} \xi_{x_i}(a, b), \quad a, b \in X \quad (5)$$

4. Полученное отношение является отношением α - толерантности на множестве X .

5. Построить транзитивное замыкание отношения меры сходства образцов данных на множестве X

$$R_\xi^1 = R_\xi$$

По $q = \overline{2, N}$, строим цикл

$$R_{\xi}^q = R_{\xi}^{q-1} \circ R_{\xi}$$

$$R_{\xi}^{|X|} = R_{\xi}^N, r_{\xi}^{|X|} = r_{\xi}^N$$

Построенное отношение $R_{\xi}^{|X|}$ есть отношение α - квазиэквивалентности.

Построим для отношения α - квазиэквивалентности шкалу α - квазиэквивалентности как множество различных элементов отношения $R_{\xi}^{|X|}$.

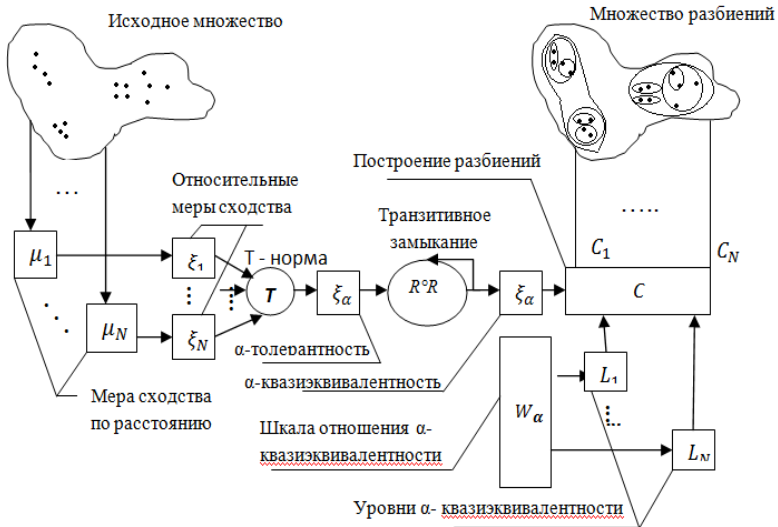


Рис.1. Этапы выполнения алгоритма

Конец алгоритма. [1]

Данный алгоритм проиллюстрирован на рис. 1.

Кластеризация «голубых фишек» Нью-Йоркской фондовой биржи.

Проведём кластеризацию наиболее ликвидных акций Нью-Йоркской фондовой биржи, с помощью α - квазиэквивалентности.

Для этого были взяты колебание акций с 01.02.2010 по 01.04.2010 [3]. В выборку вошли 10 корпораций: Apple, Intel Corp, IBM, Cisco, Google, Microsoft, YAHOO, XEROX, Lexmark International, Computer Sciences Corp.

Матрица эвклидовых расстояниями между точками будет выглядеть следующим образом (табл. 1).

Таблица 1

Матрица эвклидовых расстояний

	APL	CSC	CSO	GOG	IBM	LEX	MIC	XER	YAH	INT
APL	0	26,532	12,4006 6	24,0060 8	12,7636 6	24,8845 4	26,5222 8	24,5281 3	14,1607 1	20,9845 6
CSC	26,532	0	26,9348 2	15,5718 6	25,4536 2	14,7301 6	36,1713 7	13,818	26,3179 5	14,5271 9
CSO	12,4006 6	26,9348	0	22,6515 3	12,2930 9	25,7311 4	24,2986 1	23,5175 5	14,6246 8	20,7521 3
GO G	24,0060 8	57186	22,6515 3	0	22,0744 5	16,4221 9	34,3717 3	13,3972 9	24,1627 6	12,7258 9
IBM	12,7636 6	25,4536 2	12,2930 9	22,0744 5	0	24,295	27,8342 5	22,5459 9	15,6679 5	21,7177 7
LEX	24,8845 4	14,7301 6	25,7311 4	16,4221 9	24,295	0	35,8777 1	15,5995 4	25,1770 7	15,7447 5
MIC	26,5222 8	36,1713 7	24,2986 1	34,3717 3	27,8342 5	35,8777 1	0	33,7077 4	24,6882 9	31,9138 8
XER	24,5281 3	18	23,5175 5	13,3972 9	22,5459 9	15,5995 4	33,7077 4	0	24,7434 9	13,9989 1
YA H	14,1607 1	26,3179 5	14,6246 8	24,1627 6	15,6679 5	25,1770 7	24,6882 9	24,7434 9	0	20,0076 3
INT	20,9845 6	14,5271 9	20,7521 3	12,7258 9	21,7177 7	15,7447 5	31,9138 8	13,9989 1	20,0076 3	0

Матрица отношения сходства образцов данных на множестве X выглядит следующим образом (табл. 2).

Таблица 2 Матрица отношения сходства образцов данных на множестве X

	APL	CSC	CSO	GOG	IBM	LEX	MIC	XER	YAH	INT
APL	1	0	0,532615	0,095203	0,518934	0,062093	0,000366	0,075527	0,461937	0,209085
CSC	0	1	0	0,546957	0,085529	0,589434	0	0,590064	0	0,5448
CSO	0,532615	0	1	0,159024	0,543599	0,044689	0,097873	0,126872	0,444308	0,229543
GOG	0,095203	0,546957	0,159024	1	0,206932	0,522218	0	0,602546	0,081891	0,601243
IBM	0,518934	0,085529	0,543599	0,206932	1	0,127155	0	0,189991	0,404667	0,219747
LEX	0,062093	0,589434	0,044689	0,522218	0,127155	1	0	0,537212	0,04335	0,506649
MIC	0,000366	0	0,097873	0	0	0	1	0	0,061922	0
XER	0,075527	0,590064	0,126872	0,602546	0,189991	0,537212	0	1	0,059825	0,561353
YAH	0,461937	0	0,444308	0,081891	0,404667	0,04335	0,061922	0,059825	1	0,239773
INT	0,209085	0,5448	0,229543	0,601243	0,219747	0,506649	0	0,561353	0,239773	1

Матрица транзитивного замыкания отношения сходства данных на множестве X (матрица отношения α - квазиэквивалентности) выглядит следующим образом (табл. 3).

Таблица 3. Матрица α - квазиэквивалентности

	APL	CSC	CSO	GOG	IBM	LEX	MIC	XER	YAH	INT
APL	1	0,239773	0,532615	0,239773	0,532615	0,239773	0,097873	0,239773	0,461937	0,239773
CSC	0,239773	1	0,239773	0,590064	0,239773	0,589434	0,097873	0,590064	0,239773	0,590064
CSO	0,532615	0,239773	1	0,239773	0,543599	0,239773	0,097873	0,239773	0,461937	0,239773
GOG	0,239773	0,590064	0,239773	1	0,239773	0,589434	0,097873	0,602546	0,239773	0,601243
IBM	0,532615	0,239773	0,543599	0,239773	1	0,239773	0,097873	0,239773	0,461937	0,239773
LEX	0,239773	0,589434	0,239773	0,589434	0,239773	1	0,097873	0,589434	0,239773	0,589434
MIC	0,097873	0,097873	0,097873	0,097873	0,097873	0,097873	1	0,097873	0,097873	0,097873
XER	0,239773	0,590064	0,239773	0,602546	0,239773	0,589434	0,097873	1	0,239773	0,601243
YAH	0,461937	0,239773	0,461937	0,239773	0,461937	0,239773	0,097873	0,239773	1	0,239773
INT	0,239773	0,590064	0,239773	0,601243	0,239773	0,589434	0,097873	0,601243	0,239773	1

На основе последней таблицы получаем шкалу отношения α - квазиэквивалентности и соответствующие классы эквивалентности (табл. 4).

Таблица 4. Соотношение шкалы α - квазиэквивалентности и соответствующих классов эквивалентности

Уровень отношения α - квазиэквивалентности (Количество классов эквивалентности)	Разбиение множества ценных бумаг
1/(10)	{APL}, {CSC}, {CSO}, {GOG}, {IBM}, {INT}, {LEX}, {MIC}, {XER}, {YAH}
0,6025/(2)	{XER, GOG }
0,6012/3	{ INT , GOG, XER }
0,59/4	{ INT , GOG, CSC, XER }
0,5894/5	{ INT , GOG, CSC, XER, LEX }
0,5436/2	{IBM, CSO}
0,5326/3	{APL, IBM, CSO}
0,4619/4	{APL, IBM, CSO, YAH}
0,2398/9	{APL, CSC, GOG ,LEX, XER, INT, IBM, CSO, YAH}
0,0979/10	{APL, CSC, GOG ,LEX, XER, INT, IBM, CSO, YAH, MIC}

На выходе мы получили разбиение на кластеры в зависимости от α - квазиэквивалентности. Проанализировав полученные результаты, мы можем разбить наши ценные бумаги на 2 группы. В первый кластер вошли акции корпораций Intel Corp, Google, XEROX, Lexmark International, Computer Sciences Corp. Во вторую: Apple, IBM, Cisco, YAHOO. Ценные бумаги Microsoft лучше не включать в группы, так, как уровень α - квазиэквивалентности с этой ЦБ и всеми остальными составляет 0,0979.

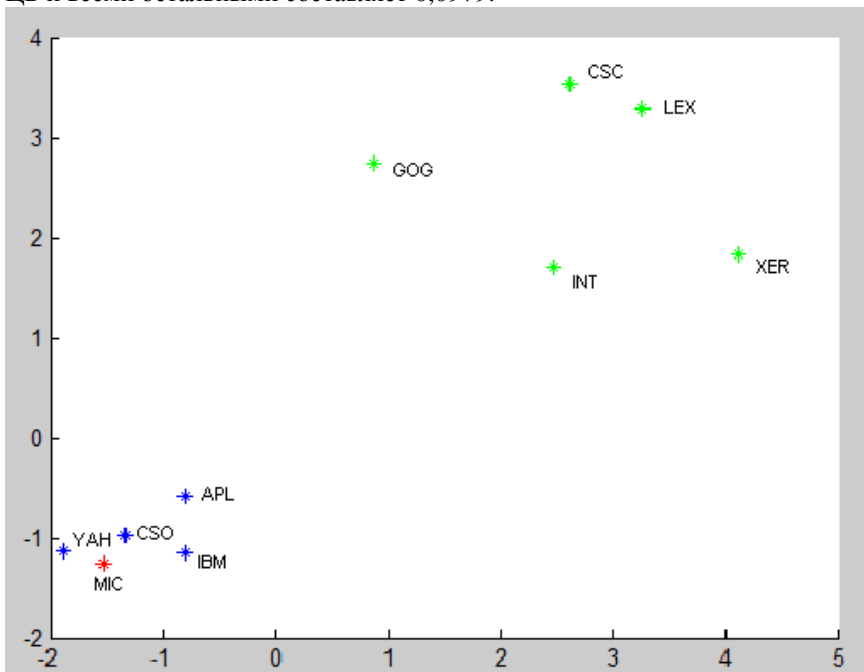


Рис.2. Кластеризация ценных бумаг методом fuzzy C-Means (разбитие на 3 кластера)

Выше приведён результат кластеризации ценных бумаг методом нечётких k- средних на 3 кластера. В 1-ый кластер вошли акции GOG, XER,INT, CSC,LEX во 2-ой кластер: APL,CSO,IBM,YAH и в 3-ий входит 1 ЦБ – MIC. Как видим разбиение на кластеры двумя разными методами нечёткой кластеризации дало нам идентичные результаты, что дает нам возможность использовать полученные группы в дальнейшем как выборки для дальнейшего прогнозирования курса акций с помощью нечётких нейронных сетей или НМГУА и формирования с них портфелей.

Выводы

Кластерный анализ включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры. В общем, каждый раз, когда необходимо классифицировать "горы" информации пригодных для дальнейшей обработки группами, кластерный анализ оказывается весьма полезным и эффективным. Кластерный анализ необходим для классификации информации, с его помощью можно определенным образом структурировать переменные и узнать, какие переменные объединяются в первую очередь, а какие следует рассматривать отдельно.

Большое достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Кроме того, кластерный анализ в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов, и позволяет рассматривать множество исходных данных практически произвольной природы. Это имеет большое значение, например, для прогнозирования конъюнктуры, когда показатели имеют разнообразный вид, что затрудняет применение традиционных эконометрических подходов.

Способом кластеризации в данном случае в соответствии с выбранным подходом является применение семейства отношений эквивалентности. Каждое из этих отношений получается при помощи перехода от отношения α - квазиэквивалентности к отношению эквивалентности в классическом смысле с использованием соответствующего уровня отношения α - квазиэквивалентности из шкалы отношения α - квазиэквивалентности. Те образцы данных, которые в соответствии с отношением α - квазиэквивалентности имеют сходство, превышающее указанный уровень, являются эквивалентными, остальные — неэквивалентными.

Описанная методика позволяет оптимально решить сразу две важнейшие проблемы: разбиение множества ценных бумаг на отдельные однородные группы, а также выявление факторов воздействия внешней среды, влияющие на данные группы с последующим пребыванием факторных весов. Это позволяет избежать сортировки компаний исключительно по отраслям (например, с использованием сектор - факторов). Полученные группы являются входными выборками в нечеткие нейронные сети, или для НМГУА для получения прогноза данных акций.

Литература

1. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. К.: Слово, 2008. - 344 с.
2. Севриновский В.Д. Методика формирования инвестиционного портфеля с использованием факторного и кластерного анализа. - М., 1998.
3. Сайт Нью-Йоркской фондовой биржи <http://www.nyse.com>