

Визуализация многомерных временных рядов и ее применение в принятии решений

Киевский национальный университет строительства и архитектуры

Рассматриваются методы визуализации многомерных временных рядов. Предложен метод и алгоритм представления многомерного временного ряда и его отдельных участков в виде иерархической структуры, получаемой при помощи кластерного анализа. Приведен пример.

Введение

В последнее время резко усилился интерес к задачам, формулирование и решение которых выполняется методами и в рамках Интернет-математики, одним из основных направлений которой является *визуальный компьютеринг* (ВК). Главной задачей ВК является интеллектуальный анализ данных, в состав которого включают, в частности, Soft Computing, нейронные сети, нечеткую математику и логику, гранулярный компьютеринг и др. В свою очередь, гранулярный компьютеринг понимают как "... современную парадигму, методологию и методику информационного анализа неочевидно структурированных динамических систем («нечетко» структурированных систем) с изменяющейся в процессе функционирования структурой"[1].

Особое место в этом смысле занимают многомерные временные ряды (МВР), в частности, временные ряды (ВР), представляющие изменение во времени экономических и финансовых параметров, которые не целесообразно рассматриваться семантически раздельно. Например, рынки ценных бумаг (ЦБ) характеризуются стоимостью, доходностью и риском ЦБ, к тому же эти параметры, в частности, доходность и риск, как правило, не известны точно и решение приходится принимать в условиях неопределенности [2,3].

Многие эксперты сходятся во мнении, что большинство решений приходится принимать не дожидаясь точного математического анализа МВР, располагая лишь характеристиками, получаемыми путем визуального анализа. Кроме того, спецификой современного управления практически во всех сферах является необходимость обработки

сверхбольших объемов информации, изменяющихся в режиме реального времени, что невозможно выполнить с помощью стандартных методов на современных ПК. Отметим также, что современное управление невозможно без обработки сверхбольших массивов визуальной информации, ВК должен позволить преодолеть этот «кризис данных».

Современное состояние исследований

Прежде всего отметим, что наибольшая рациональность использования ВК достигается в среде Интернет-ориентированных задач, хотя можно указать целый ряд задач управления в условиях неопределенности, где визуализация не только дает существенный эффект, но в ряде случаев является лучшим средством приближенного решения задачи. Специфика решения задач при визуализации взаимодействия пользователей Интернета состоит в том, что исследователи стараются подобрать такую форму визуализации, чтобы увидеть аналогии с хорошо изученными системами, в частности, биологическими или др. системами [1], поведение которых достаточно хорошо исследовано или просто известно. Это позволяет применить т.н. визуальную аналитику [4,5] для дальнейшего конкретного выявления общесистемных закономерностей и различий.

В работе [6] приведено описание системы Visual Analytics, включающей методы визуальных рассуждений, визуального представления данных, текстов и объяснений, когнитивные и эвристические алгоритмы визуального интерактивного анализа, разработку и использование визуального моделирования, визуальное представление знаний и т.д. Отмечается, что ВК позволяет эффек-

тивно пояснювати, аналізувати, моделювати і прогнозувати строєння, динаміку і зміст багатьох процесів.

В більшості робіт відзначається, що ВК тісно пов'язаний з гранулярним комп'ютерингом (ГрК). Вперше питання гранулювання нечіткої інформації і способи її обробки поставлені в роботі Л.Заде [7], в якій *гранула* визначена як група об'єктів (або точок), які представлені *вместе* на основі нерозличимості, схожості, близькості. Концепція *універсального обмеження* [8] забезпечує основу для класифікації нечітких гранул (Γ). В теорії нечіткості Γ розглянута як група точок, характеризується універсальним обмеженням, тип Γ визначений типом обмежень. Питання формування інформаційних гранул тісно пов'язані з теорією грубих множин (ΓM). В роботі [9] показано, що ΓM може розглядатися як чітке множество *с грубим описанням*.

В якості критеріїв грануляції використовують різні формальні і змістові критерії, наприклад, оптимізаційний принцип мінімакса, критерії інформаційної важливості, точності прогнозування і т.д.

Постановка задачі.

Одним з найбільш вивчених методів ВК є ієрархічний кластерний аналіз [10]. Під кластером зазвичай розуміється частина даних (в типовому випадку – підмножина об'єктів або підмножина змінних, або підмножина об'єктів, характеризується підмножиною змінних), яка виділяється з решти частини наявності певної однорідності її елементів або її відсутності [11]. В найпростішому випадку мова йде про схожість елементів, в ідеальному випадку – про збігаючіся значення основних змінних або іншого роду *близькості*, виражаємої геометричною близькістю відповідних об'єктів. ГрК в цьому випадку розглядається як робота з грануло-кластером, що дозволяє дати багаторівневе «нечітке» описання і пояснення неочевидно структурованої динамічної системи, якою є ВР, ієрархічна структура якої змінюється в процесі функціонування.

Возможності кластер-аналізу в візуальній підтримці прийняття рішень раціонально використовувати зазвичай для оцінки аномальності окремих частин (вікон) МВР, що дозволяє використовувати більш різноманітну, зрозумілу і формалізовану систему правил. Сюди слід включати такі задачі підтримки прийняття рішень, як:

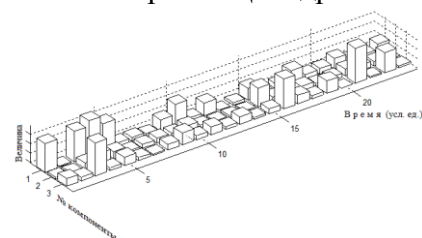
- структурізація даних;
- виявлення груп об'єктів, до яких застосовні однакові критерії;
- виявлення і аналіз структури взаємодії основних підсистем;
- візуалізація структури МВР.

Візуалізація залишається одним з основних засобів підтримки прийняття рішень, стимулюючих інтуїцію. Після того як основні кластери встановлені, їх взаємодія і розвиток значно легше візуалізувати, ніж в початковій масі даних.

Алгоритм рішення

Предполагає. МВР має свою ієрархічну структуру, яка *нормально* не змінюється суттєво або змінюється в контролюємуєму межах, якщо ВР не містить *аномалій* на окремих ділянках його дослідження. В якості *аномалій* розглядаються зміни структури на окремій ділянці, відхилення в величинах (по порівнянню з стандартними), зміни співвідношень (відстаней) між величинами, зміна знаку трендів на окремих ділянках і др.

Розглядаємуєму МВР наведено на рис.1, в найпростішому випадку маємо масив $\left((x_1^{(1)} x_2^{(1)} x_3^{(1)}); \dots (x_1^{(n)} x_2^{(n)} x_3^{(n)}) \right)$, де кожен рядок матриці – вектор $x^{(i)} = \{ x_1^{(i)} x_2^{(i)} x_3^{(i)} \}$ – значення МВР в момент часу $i, i=1, n$. В економічних задачах даний ВР може інтерпретуватися в координатах «дохідність-ціна-ризик» ЦБ і др.



а)

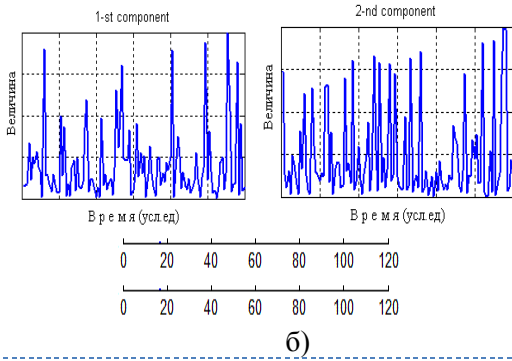
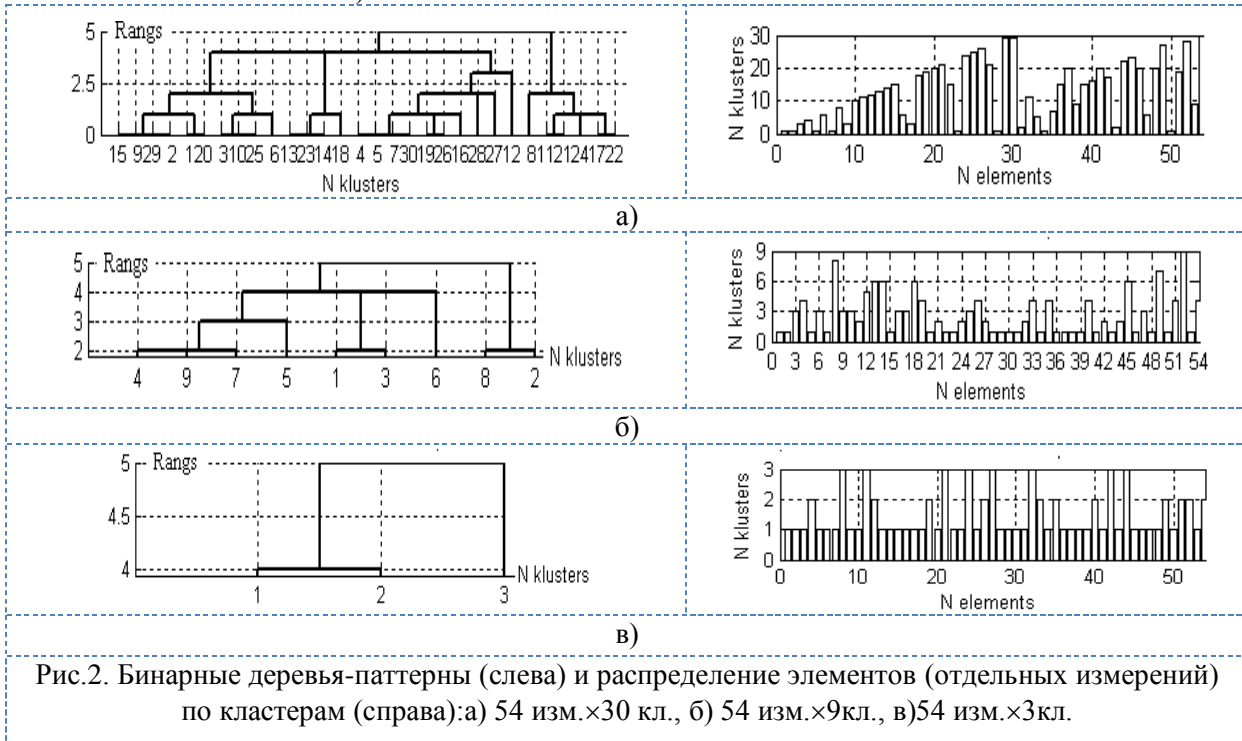


Рис.1. Пример 3-х мерного ВР: а) общий вид, б) изменение компонент во времени.

На рис. 2 приведены бинарные деревья-паттерны и распределение элементов (отдельных измерений) по кластерам для следующих случаев: а) 54 изм. × 30 кл., б) 54 изм. × 9 кл., в) 54 изм. × 3 кл.



Ниже представлен фрагмент матрицы размерностью 3 × 54 (рис 2.в).

$$\begin{pmatrix} \\ \\ \vdots \\ \end{pmatrix}_{46}^{54} \rightarrow \begin{pmatrix} 0.3032 & 0.3336 & 0.0648 \\ 0.1024 & 0.1251 & 0.0435 \\ 0.2266 & 0.1820 & 0.8226 \\ 1.1369 & 1.0525 & 1.1018 \\ \vdots & \vdots & \vdots \\ 1.7479 & 1.6033 & 0.0394 \\ 0.7340 & 0.6871 & 0.1771 \\ 0.9660 & 0.2029 & 1.2354 \end{pmatrix}_{1}^{54}$$

Специфика деревьев в том, что они позволяют с различной степенью детализации исследовать объект, в частности, особенности иерархической структуры, что имеет при визуальном компьютеринге достаточно большое значение. Наиболее грубая гранула – рис. 2-в имеет матрицу 3×5, что резко упрощает процесс восприятия объекта и позволяет принять решение более просто, естественно, с меньшей точностью. Выбор большого (по количе-

ству элементов) паттерна практически не оправдан, наиболее рациональным является паттерн, совпадающий по количеству элементов с исследуемой частью ВР.

Ниже приведены исследования последовательных участков МВР, представленных матрицами 9×3 – рис.3. Нормы, математическое ожидание (МО) и дисперсии этих участков МВР таковы:

НОРМЫ=[5.1054 8.9288 5.8364 3.3652 6.4912 5.3247],

МО=[0.5188 1.2070 0.7504 0.5352 0.8706 0.6766],

ДИСПЕРСИИ=[0.4329 0.2512 0.3654 0.0897 0.2271 0.2960],

для всей рассматриваемой (объединенной) части МВР 3×54:F -норма = 14.8884, МО=0.7598, дисперсия =0.1098.

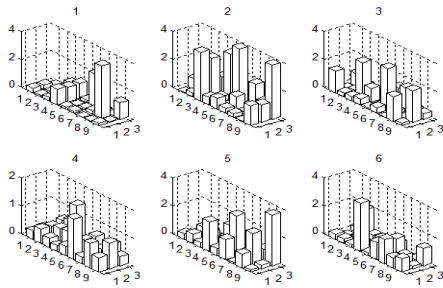


Рис. 3. Последовательные участки ВР

Структурные характеристики последовательных участков МВР определены на основании дендрограмм (рис.6), полученных по нижеприведенному алгоритму:

```

y=pdist(Y(x_{j+1}...x_{j+9}, :,:))
          9 измерений
z=linkage(y,'complete')
subplot(231),
[H T PERM1]=dendrogram(fix(z),m)
set(H,'LineWidth',2)
grid on
title('1-st part')
    
```

На рис. 4 приводятся дендрограммы 6-ти последовательных участков МВР.

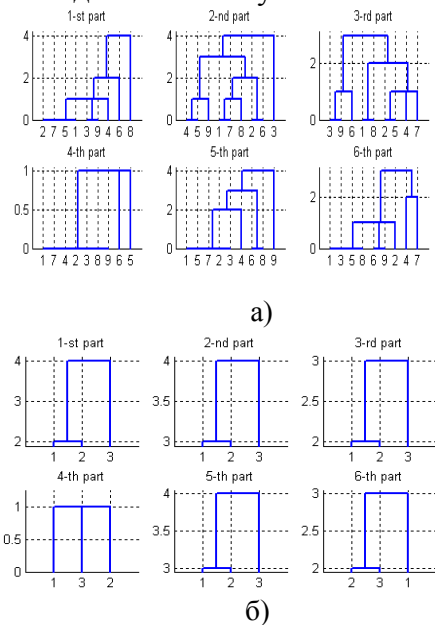


Рис.4. Дендрограммы 6 последовательных участков МВР: а- 1изм × 1 кл., б) 3 изм × 1 кл

На рис.5 приведены бинарные матрицы для дендрограмм, представленных на рис.4-б, вычислены параметры иерархических структур: 2-адическое число и 2-адическая размерность.

Последовательные участки МВР

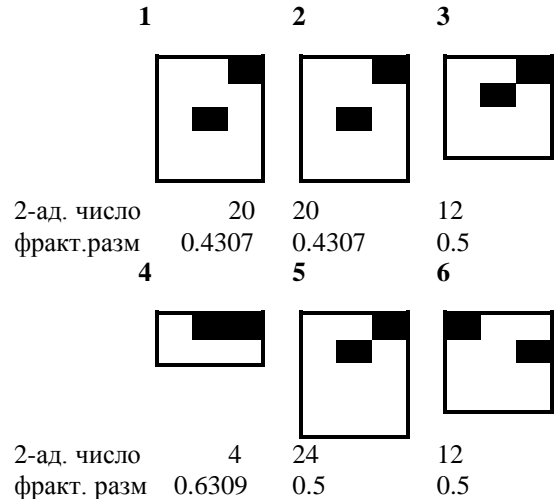


Рис. 5. Бинарные матрицы для дендрограмм, представленных на рис.4-б.

Если выбрать в качестве паттерна участок №1 МВР, то на основании рис. 4 и рис.5 можно утверждать, что участки с номерами 2 и 5 наиболее близки к паттерну, участок №4 наименее близок, участки с номерами 3 и 6 скорее близки и занимают промежуточное положение между паттерном и наиболее удаленным участком. Структурное нивелирование свойств имеет место при неоправданном увеличении размера паттерна. Дендрограммы для паттерна, представляющего собой объединение всех участков, представлены на рис.6, а – кластеризация 54 изм × 9 кл., и б - 54 изм × 3 кл. Такая форма кластеризации обусловлена необходимостью визуального сравнения структурных свойств паттерна и последовательных участков МВР (рис. 4 и рис. 6).

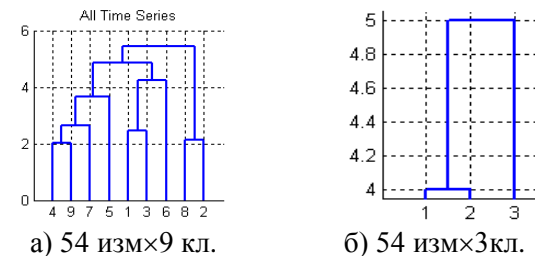


Рис.6. Дендрограммы 6 последовательных участков МВР.

Выполняя визуальное и текстуальное сравнение можно видеть, что выводы, полученные в предыдущем случае, сохраняются и в данном случае.

Иерархическая структура паттерна и тестовых участков МВР, полученная методом наиболее близкого соседа, приведена на рис. 7 и 8, алгоритм кластеризации приведен ниже.

```

y=pdist(Y(9 измерений ;:))
z=linkage(y,'single')
subplot(231),
[H T PERM1]=dendrogram(fix(z),m)
set(H,'LineWidth',2)
grid on
title('1-st part')

```

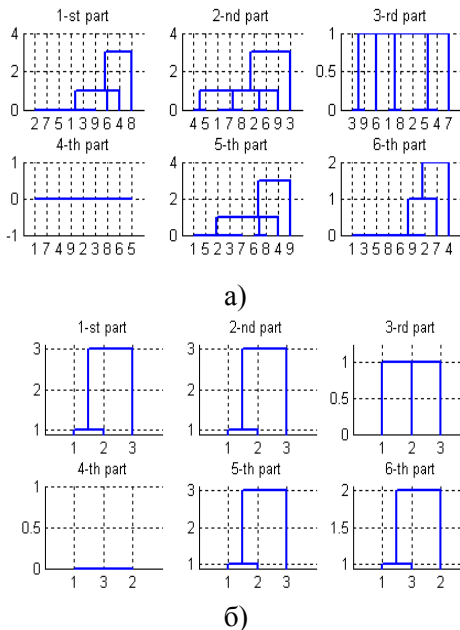


Рис. 7. Дендрограммы 6 последовательных участков МВР: а- 1изм × 1 кл., б) 3 изм × 1 кл.

Параметры иерархических структур: 2- адическое число и 2-адическая размерность.

Последовательные участки МВР

	1	2	3
2-ад. число	10	10	4
фракт. разм	0.5	0.5	0.6309
	4	5	6
2-ад. число	0	10	6
фракт. разм	inf	0.5	0.6309

Рис. 8. Бинарные матрицы для дендрограмм, представленных на рис.4-6.

На рис. 9 приводится дендрограмма паттерна МВР 54 изм×3 кл, ее бинарная матрица, для паттерна 2-адическое число равно 6, 2-адическая размерность - 0.6309. Сравнивая рис. 8 и рис. 9, отметим, что участок №4 однозначно идентифицируется как аномальный, наиболее близкий к паттерну участок №6, близкие - участки №3 и №2.

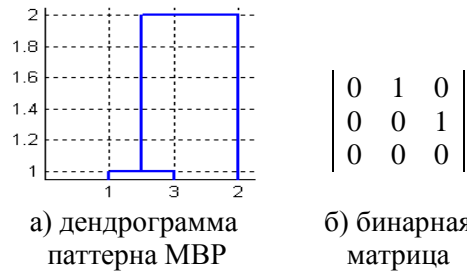


Рис. 9. Характеристики паттерна МВР 54изм×3 кл

Совместное использование 2-х способов кластеризации требует отдельного обоснования, однако одинаково однозначная идентификация аномального участка является успешным решением поставленной задачи. Естественно, в зависимости от смысла задачи необходим дополнительный анализ причин данного факта.

Выводы

1. В современных условиях значительную часть задач экономики, менеджмента и др. приходится формулировать и решать методами Интернет-математики, одним из основных направлений которой является визуальный компьютеринг. Главной задачей ВК есть интеллектуальный анализ данных, в частности, анализ МВР. Специфика задач состоит в том, что большинство решений приходится принимать, не дожидаясь точного математического анализа МВР, располагая лишь характеристиками, получаемыми путем визуального анализа. Кроме того, сложностью современного управления практически во всех сферах является необходимость обработки сверхбольших объемов информации, изменяющихся в режиме реального времени.

2. Предложены метод и алгоритм представления МВР и его отдельных участков в виде иерархической структуры, получаемой при помощи кластерного анализа. Предложены характеристики иерархической структуры – максимальный ранг бинарного дерева, фрактальное (2-адическое) число и фрактальная (2-адическая) размерность, позволяющие

оценить близость МВР на основании данных характеристик и способствовать принятию решения. Приведенные примеры показали эффективность предложенного метода.

Список литературы

1. Давыдов А.А. О компьютерной теории социальных агентов//Социол.исслед. 2006, №2, С.19-28.
2. Трофимов Г.Ю. Экономика и нейронаука - на пути синтеза//Экономика и мат. методы. 2006, Т.42, № 4, С. 3–16.
3. Браун С. Трехмерные изображения//В мире науки, 2007, №9. Интернет-ресурс:
<http://www.sciam.ru/2007/9/innovation.shtml>.
4. Fabien B. Analytical System Dynamics: Modeling and Simulation. N.Y.: Springer, 2008.
5. Les Z., Les M. Shape Understanding System: The First Steps toward the Visual Thinking Machines. N.Y.: Springer, 2008.
6. Интернет-ресурс:
<http://graphics.cs.brown.edu/research/cave/brain.jpg>
7. Zadeh L.A. Toward a Generalized Theory of Uncertainty// Information Sciences –

Informatics and Computer Science.–2005.– Vol.172. – P.1-40.

8. Aja-Fern´andez S., Alberola-L´opez C. Fuzzy Granules as a BasicWord Representation for Computing with Words. SPECOM'2004: 9th Conference Speech and Computer St. Petersburg, Russia September 20-22, 2004.ISCA Archive. Интернет-ресурс
[/http://www.isca-speech.org/archive](http://www.isca-speech.org/archive)

9. Pawlak, Z., Skowron, A.: Rough membership functions. In: Advances in the Dempster Shafer Theory of Evidence, (R.R. Yaeger, M. Fedrizzi and J. Kasprzyk eds.), John Wiley&Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1994.

10. Минаев Ю.Н., Филимонова О.Ю., Минаева Ю.И. Иерархическая кластеризация нечетких данных. Электрон. моделир. 2012, т.34, №4.-с.3-22

11. Мінаєва Ю.І. Інформаційна інтелектуальна технологія формування ПЦП в умовах невизначеності: автореф. дис. на здобуття наук. ступеня кандидата техн.наук: спец.05.13.06 "Інформаційні технології"/Ю.І.Мінаєва. - Київ, 2012. - 20 с.