

Визуализация точек многомерного информационного текстового массива на упругой карте для оценки кластерной структуры данных

А.Е. Бондарев^{1,А}

Институт прикладной математики им. М.В. Келдыша Российской академии наук

¹ ORCID: 0000-0003-3681-5212, bond@keldysh.ru

Аннотация

Представлены результаты вычислительных экспериментов по отображению точек исходного многомерного информационного массива на развертку упругой карты для оценки взаиморасположения областей смысловой близости с целью совершенствования обработки текстовой информации. Упругие карты рассматриваются в роли инструмента обеспечения аналитической работы с текстовой информацией. Как показывают предыдущие работы, для получения нужных расстояний, соответствующих кластерной картине изучаемого многомерного объема, необходимо использовать расстояния на упругой карте, которая и отражает кластерный портрет исследуемого многомерного объема данных. В работе представлены полученные таким способом кластерные структуры точек изучаемого многомерного объема на развертке упругой карты в плоскости двух первых главных компонент. Представлен анализ взаиморасположения кластеров различных конфигураций на разные моменты времени.

Ключевые слова: Многомерные текстовые данные, кластерная структура, упругие карты, анализ взаиморасположения кластеров.

1. Введение

Данная работа представляет результаты проведения вычислительных экспериментов по кластерному анализу для многомерных информационных массивов. Многомерные информационные массивы представляют собой текстовую информацию, являясь в цифровом виде частотами совместного употребления слов из различных частей речи (например, существительное + прилагательное). Для получения такого многомерного массива проводится анализ текстовых коллекций. В результате, для 300-мерного массива, построенного таким образом, мы рассматриваем 300 точек в пространстве из 300 измерений. Данная работа является продолжением работ [19-26] и во многом использует алгоритмы, результаты и наработки из предыдущих работ.

На современном этапе чрезвычайно актуальной задачей становится изучение и анализ многомерных объемов данных. Анализ многомерных данных является актуальной проблемой достаточно давно. Для подобного изучения разрабатывались методы анализа данных, методы визуальной аналитики. Применение этих методов дает возможность на практике узнать структуру исследуемого объема многомерных данных, увидеть его кластерную картину, определить области сгущения данных (кластеры), и т.д.

Изучение многомерных данных сформировало новую междисциплинарную область исследований, как визуальная аналитика (Visual Analytics). Визуальная аналитика – это комплекс методов и подходов, направленных на проведение анализа и визуального представления многомерных данных независимо от природы их происхождения. Основные концепции визуальной аналитики представлены в первых работах по данной дисциплине [1-4]. Подходы и методы визуальной аналитики и позволяют решать ряд

практических задач исследования многомерных данных. К таким задачам можно отнести задачи классификации данных, обнаружения кластеров, выявление ключевых определяющих параметров, установление взаимосвязей между ключевыми параметрами и т.д.

Для отображения многомерного объема данных во вложенные в исходный объем многообразия меньшей размерности были разработаны подходы самоорганизующихся карт [5-7]. К подобным картам относятся и так называемые упругие карты (Elastic Maps). Теория упругих карт была разработана А. Горбанем и А. Зиновьевым вместе с коллегами и достаточно подробно описана в работах [8-14]. Упругие карты нашли довольно широкое применение в практическом анализе многомерных данных, причем в самых разных областях – анализ экономических данных, анализ полезных ископаемых, анализ медицинских данных и т.п. Примеры исследований представлены в работах [15-18]. Важнейшим свойством упругих карт является способность успешно использовать их для любых многомерных данных независимо от природы их происхождения. Особую эффективность в решении задач кластерного анализа многомерных данных упругие карты приобретают при совместном использовании с методом главных компонент (PCA). Отображение упругой карты и ее развертки в пространстве, образованном первыми тремя главными компонентами, в задачах кластеризации и классификации позволяет более точно и четко определять кластерную структуру изучаемых многомерных объемов данных. На рисунке 1 приведен пример развертки упругой карты на плоскость, расположенную в пространстве главных компонент с раскраской по плотности данных. Подобное представление позволяет определить кластерную структуру изучаемого объема многомерных данных без применения специальных алгоритмов кластеризации.

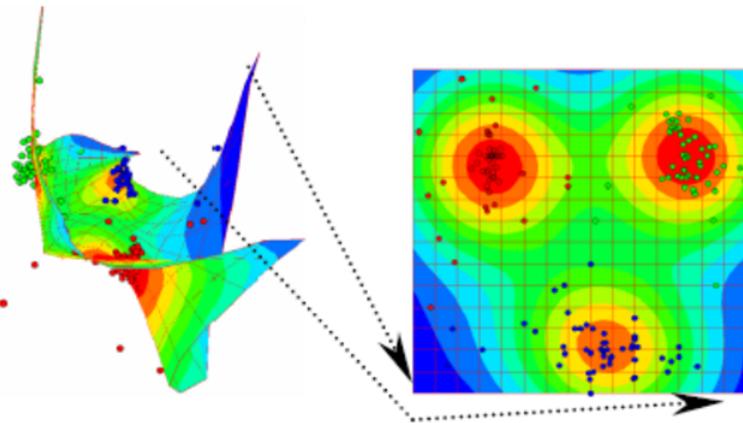


Рис.1. Пример упругой карты и ее развертки на плоскость с раскраской по плотности данных.

Важное значение имеет применение построения упругих карт к задачам анализа многомерных текстовых данных. Это направление появилось из необходимости решения практических задач кластерного анализа и становится все более актуальным. Применение построения упругих карт к задачам анализа многомерных текстовых данных впервые было реализовано на практике А.Е. Бондаревым и В.А. Галактионовым в ИПМ им. М.В. Келдыша РАН. В качестве изучаемых объемов многомерных данных использовались массивы частот совместного употребления различных частей речи, получаемых из текстовых коллекций. Результаты исследований подробно представлены в работах [19-26]. Однако в процессе решения задач кластерного анализа многомерных информационных объемов выяснилось одно крайне важное обстоятельство. Дело в том, что получаемая в численных экспериментах картина распределения кластеров в многомерном объеме текстовых данных существует лишь на развертке упругих карт в

пространстве первых двух главных компонент. Определение конкретных внутрикластерных и межкластерных расстояний должно вестись на цифровых данных. Таковыми данными являются координаты точек многомерного объема в многомерном пространстве. Однако, как показывают работы [24-26], определение внутрикластерных и межкластерных расстояний на основе исходных координат точек многомерного пространства не дает надежных результатов. Для получения нужных расстояний необходимо использовать расстояния на упругой карте, которая и отражает кластерный портрет исследуемого многомерного объема данных. Именно этому направлению исследований посвящена данная работа.

2. Упругие карты и применение метода главных компонент (PCA) в задачах визуального анализа многомерных текстовых информационных массивов

Упругие карты являются логическим развитием карт Кохонена. Идеология и алгоритмы реализации построения упругих карт подробно представлены в работах [8 - 14]. Подобная карта представляет собой систему упругих пружин, вложенную в многомерное пространство данных. Метод упругих карт формулируется как оптимизационная задача, предполагающая оптимизацию заданного функционала от взаимного расположения карты и данных.

Согласно [8 - 14] основой для построения упругой карты является двумерная прямоугольная сетка G ,ложенная в многомерное пространство, которая аппроксимирует данные и обладает регулируемыми свойствами упругости по отношению к растяжению и изгибу. Расположение узлов сетки ищется в результате решения оптимизационной задачи на нахождение минимума функционала:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m} + \mu \frac{D_3}{m} \rightarrow \min,$$

где $|X|$ - число точек в многомерном объеме данных X ; m - число узлов сетки, λ, μ - коэффициенты упругости, отвечающие за растяжение и изогнутость сетки соответственно; D_1, D_2, D_3 - слагаемые, отвечающие за свойства сетки.

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - r^{ij}\|^2.$$

D_1 является мерой близости расположения узлов сетки к данным.

Здесь K_{ij} - подмножества точек из X , для которых узел сетки r_{ij} является ближайшим:

$$x \xrightarrow{\Pi} r^{ij}, \quad \|x - r^{ij}\|^2 \rightarrow \min, \quad K_{ij} = \left\{ x \in X, x \xrightarrow{\Pi} r^{ij} \right\},$$

Слагаемое D_2 представляет меру растянутости сетки:

$$D_2 = \sum_{ij} \|r^{ij} - r^{i,j+1}\|^2 + \sum_{ij} \|r^{ij} - r^{i+1,j}\|^2.$$

Слагаемое D_3 представляет меру изогнутости (кривизны) сетки:

$$D_3 = \sum_{ij} \|2r^{ij} - r^{i,j-1} - r^{i,j+1}\|^2 + \sum_{ij} \|2r^{ij} - r^{i-1,j} - r^{i+1,j}\|^2.$$

Варьирование параметров упругости заключается в построении упругих карт с последовательным уменьшением коэффициентов упругости, в силу чего карта становится более мягкой и гибкой, наиболее оптимальным образом подстраиваясь к точкам исходного многомерного объема данных. После построения упругую карту можно развернуть в плоскость для наблюдения кластерной структуры в изучаемом объеме данных. На развернутой плоскости можно раскраской отобразить распределение плотности данных по упругой карте. В ряде случаев подобная раскраска может оказаться весьма полезной. Особо эффективными упругие карты являются при совместном использовании с методом главных компонент (PCA). Отображение упругой карты и ее

развертки в пространстве, образованном первыми тремя главными компонентами, позволяет резко улучшить результаты, особенно в задачах кластеризации и классификации. Применение упругих карт позволяет более точно и четко определять кластерную структуру изучаемых многомерных объемов данных.

Описанный подход был многократно применен к многомерным текстовым массивам данных. Рассматривались многомерные данные, получаемые из текстовых коллекций, и представляющие частоты совместного употребления различных частей речи. Например, 300 существительных и 300 прилагательных. Прилагательные рассматривались, как измерения, а существительные – как точки в пространстве измерений. То есть, рассматривалось 300 точек в 300-мерном пространстве. На рисунке 2 представлен типичный вид упругой карты.

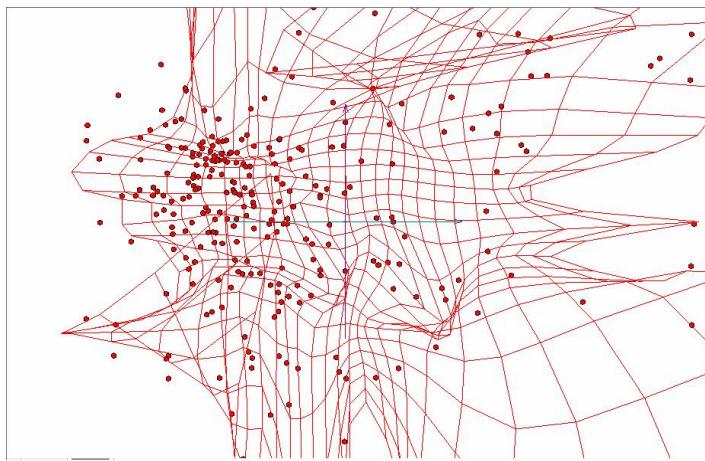


Рис.2. Типичный вид упругой карты частот совместного употребления.

3. Построение исходного набора данных и проведение тестовых расчетов взаиморасположения кластеров

Ранее в предыдущих работах [20-26] было показано, что подход оценки межкластерных расстояний с помощью построения гиперсфер в пространстве исходных данных не обеспечивает желаемого результата. Все численные исследования проводились с использованием двух типов метрик, задающих расстояние между объектами – манхэттенской метрики и косинусной метрики. Для оценки межкластерных расстояний необходимо построение других способов оценки. По результатам работ [20-26] был сделан вывод о том, что проблемы с изучением кластерных свойств связаны с тем, что исходная качественная информация о кластерах берется с развертки упругой карты, а для расчетов используются исходные координаты точек в многомерном пространстве. Было принято решение в дальнейшем использовать координаты аннотированных точек на развертке упругой карты в силу того, что они непосредственно отражают кластерную картину изучаемого объема данных.

Реальные данные были получены из текстовых корпусов новостной информации. Для получения информации из текстовых корпусов использованы процедуры, подробно описанные в [20]. Для получения необходимых многомерных массивов использовались группы сочетаний «существительное + прилагательное». Выбирались 300 существительных и 300 прилагательных. В качестве цифрового значения использовались частоты совместного употребления существительного и прилагательного в каждом изучаемом многомерном массиве. Таким образом, мы как и в прошлых экспериментах, получили возможность рассматривать прилагательные как координатные измерения в многомерном пространстве, а существительные – как точки в этом пространстве. Рассматривались 300 точек, находящихся в 300-мерном пространстве. Таким образом построено 5 многомерных массивов – за март 2005 года, за апрель 2005 года, за май 2005 года, за май 2006 года и за май 2007 года. Первые три массива позволяют проследи-

вать эволюцию кластерной структуры многомерного объема через месяц. Третий четвертый и пятый массивы позволяют прослеживать эволюцию кластерной структуры многомерного объема через год. Для всех массивов были построены упругие карты и их развертки в пространстве первых главных компонент.

4. Результаты вычислительных экспериментов

Вычислительные эксперименты включали в себя построение упругих карт и их разверток на различные моменты времени, выделение кластеров, нахождение их центров и радиусов. Напомним, что в данной задаче использовалась евклидова метрика для определения расстояний, центр каждого конкретного кластера определялся как среднее арифметическое входящих в данный кластер точек. В качестве характерного размера кластера выбиралось расстояние от центра кластера до наиболее удаленной точки кластера.

Начнем рассмотрение результатов вычислительных экспериментов с развертки упругой карты на момент времени 2005-4 (апрель 2005). Здесь в многомерном массиве данных на правом нижнем краю развертки упругой карты имеются два четко выраженных кластера, отображенных на рис.3, где представлен крупным планом фрагмент развертки, содержащий данные кластеры вместе с их вычисленными центрами.

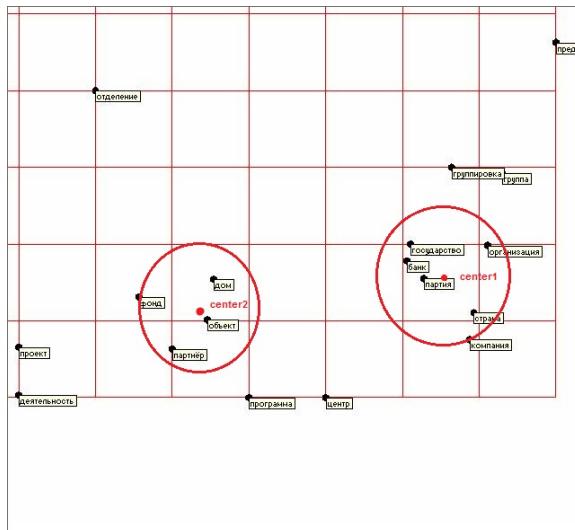


Рис. 3. Кластеры нижнего правого края для изучаемого многомерного массива на момент времени 2005-4 с центрами кластеров.

На рисунке 4 представлен фрагмент развертки упругой карты с аннотациями, соответствующая моменту времени 2005-5 (май 2005) с двумя кластерами по нижнему правому краю и центрами кластеров.

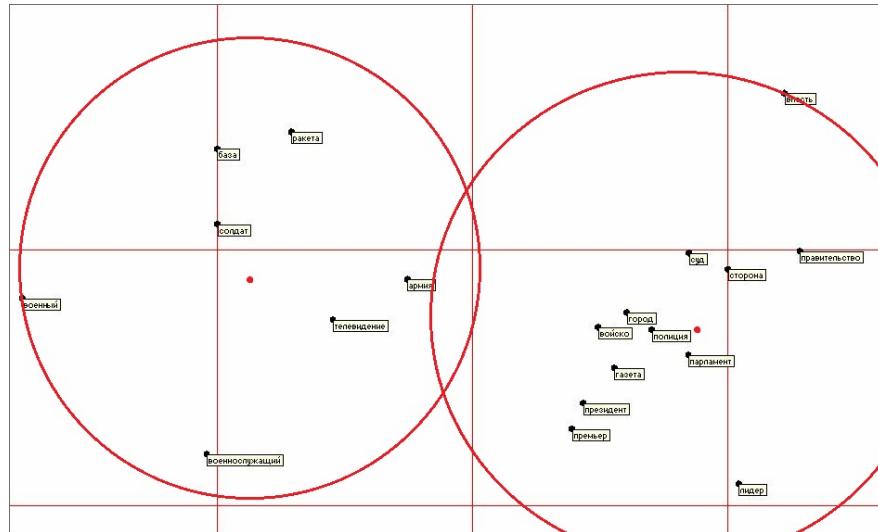


Рис. 4. Кластеры верхнего правого края для изучаемого многомерного массива на момент времени 2005-05 с центрами кластеров и радиусами.

Наиболее удаленными точками после вычисления расстояний от центра кластера являются точки «ВЛАСТЬ» и «ВОЕННЫЙ». Согласно полученным результатам в данном случае образуется маленькая область пересечения кластеров, но ни одна точка изучаемого массива многомерных данных в эту точку не попадает.

Теперь рассмотрим результаты для момента времени 2006-05. На рисунке 5 представлена развертка упругой карты с аннотациями, соответствующая моменту времени 2006-5 (май 2006).

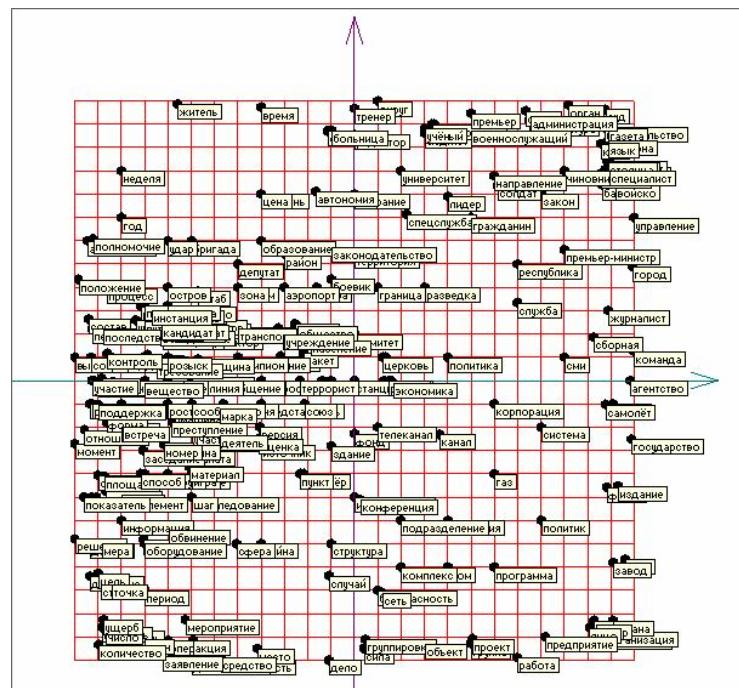


Рис.5. Аннотированная развертка упругой карты на момент времени 2006-05.

Здесь в многомерном массиве данных на правом краю в верхней и нижней частях развертки упругой карты имеются по два четко выраженных кластера. Кластеры верхней части представлены на рисунке 6.

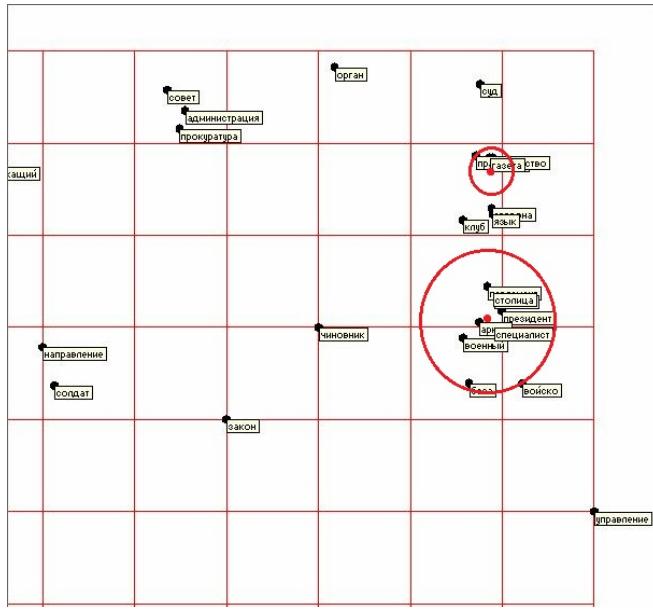


Рис. 6. Кластеры верхнего правого края для изучаемого многомерного массива на момент времени 2006-05 с центрами кластеров.

На рисунке 7 представлены кластеры нижнего правого края на момент времени 2006-5 (май 2006).

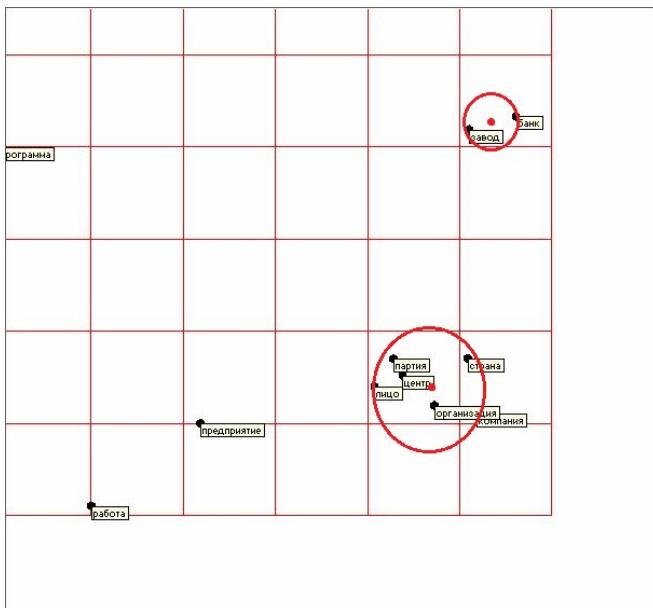


Рис. 7. Кластеры нижнего правого края для изучаемого многомерного массива на момент времени 2006-05 с центрами кластеров.

Приведенные результаты показывают, что в рассмотренных случаях пересечения кластеров и возникновения «спорных» точек нет.

Рассмотрим кластерную картину на момент времени 2007-05. Аннотированная развертка упругой карты представлена на рисунке 8.

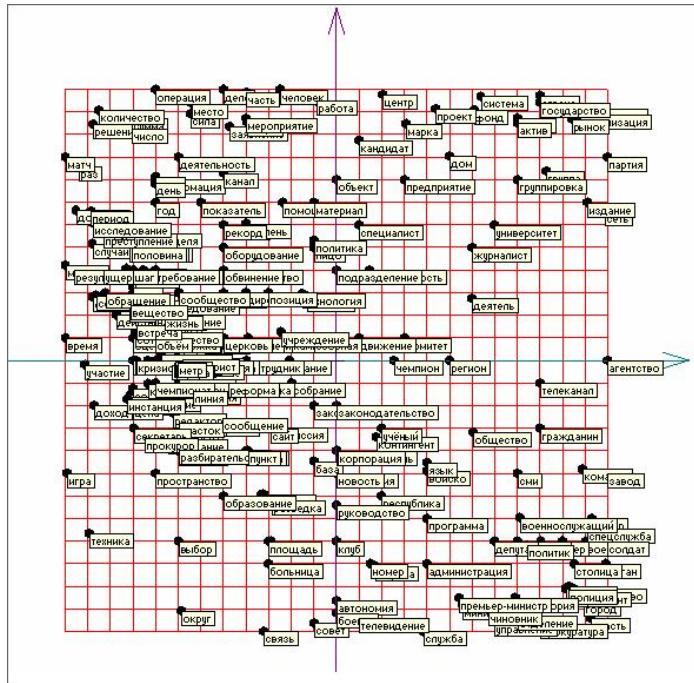


Рис.8. Аннотированная развертка упругой карты на момент времени 2007-05.

В верхней части правого угла развертки упругой карты образуются три кластера. Их вид представлен на рисунке 9.

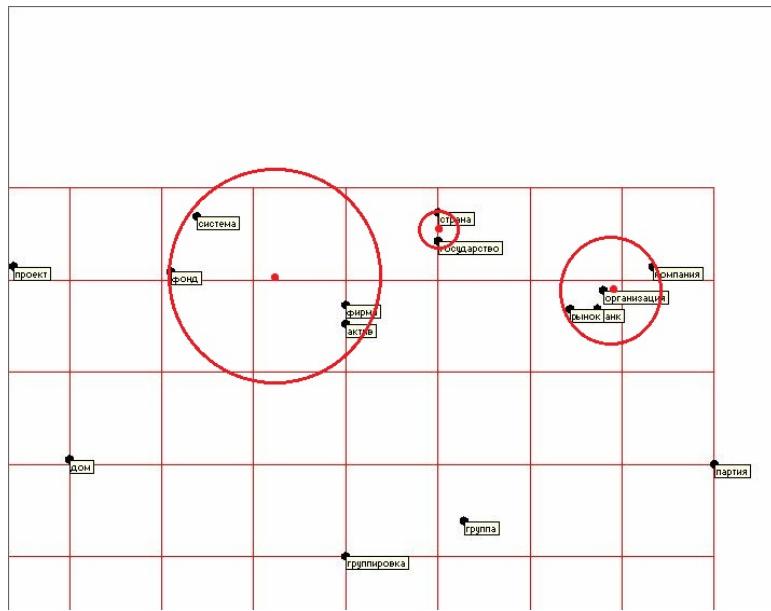


Рис.9. Кластеры верхней части правого угла развертки упругой карты на момент времени 2007-05 в виде окружностей и их центры.

Теперь представим нижнюю часть правого угла развертки упругой карты на этот же момент времени 2007-05. Здесь картина представлена на рисунке 10. Следует отметить, что в данном случае кластеры имеют более вытянутую форму, чем в предыдущих случаях.

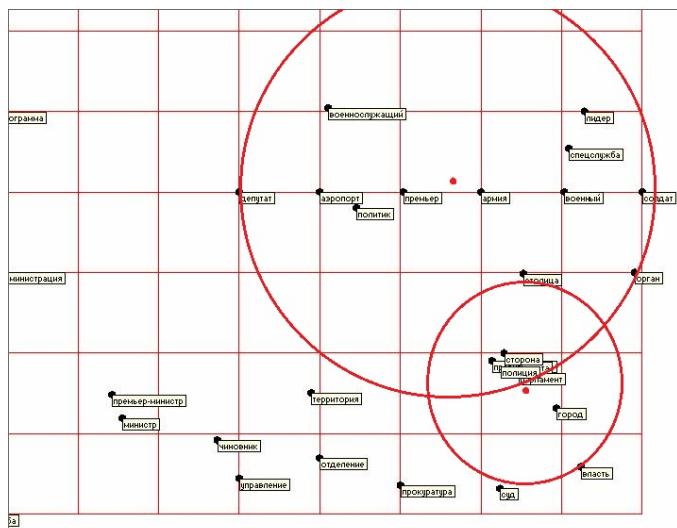


Рис.10. Кластеры нижней части правого угла развертки упругой карты на момент времени 2007-05 в виде окружностей и их центры.

Именно здесь мы сталкиваемся с той ситуацией, когда кластеры построенные по правилам, определенным нами, пересекаются. При этом пересечении в больший кластер входят «лишние» точки «СТОЛИЦА» и «ОРГАН», а также точки второго меньшего кластера. Полученный результат свидетельствует, что правила построения кластеров в подобных случаях надо регулировать так, чтобы они могли учитывать «вытянутые» кластеры, которые далеки от формы окружностей.

5. Заключение

Упругие карты, будучи представленными в развертке на плоскость, образованную двумя первыми главными компонентами, позволяют увидеть кластерную картину изучаемого многомерного объема. В качестве многомерного объема используется массив частот совместного употребления из различных частей речи (например, «прилагательное + существительное»), получаемый из текстовых коллекций. Для анализа кластерной структуры используются координаты точек многомерного объема, получаемые с развертки упругой карты. Рассмотрены взаимные расположения кластеров, возникающие на различные моменты времени. Кластеры представлены в виде окружностей. Для построения радиусов использовался принцип выбора расстояния от центра кластера до наиболее удаленной его точки.

В результате проведенных вычислительных экспериментов можно утверждать, что:

- применение упругих карт и их разверток является эффективным инструментом для аналитика при изучении многомерной текстовой информации;
 - процедуры построения кластеров и их взаиморасположения надо совершенствовать таким образом, чтобы они могли отражать «вытянутые» кластеры, проблемы пересечения кластеров, проблемы вложения кластеров;
 - необходимо на основании этих процедур разработать четкую методологию применения упругих карт и их разверток к анализу кластерной структуры многомерных информационных данных.

Список литературы

1. Thomas J. and Cook K. 2005 Illuminating the Path: Research and Development Agenda for Visual Analytics (IEEE-Press)
 2. Wong P. C., Thomas J. Visual Analytics // IEEE Computer Graphics and Applications. 2004. V. 24, N. 5. — P. 20-21.
 3. Keim D., Kohlhammer J., Ellis G. and Mansmann F. (Eds.) Mastering the Information Age – Solving Problems with Visual Analytics, Eurographics Association, 2010.

4. Kielman, J. and Thomas, J. (Guest Eds.) (2009). Special Issue: Foundations and Frontiers of Visual Analytics / Information Visualization, Volume 8, Number 4, p. 239-314.
5. T. Kohonen, Self-Organizing Maps (Third Extended Edition), New York, 2001, 501 pages.
6. Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт, Альпина Паблишер, 2001, 317 стр.
7. Кохонен Т. Самоорганизующиеся карты. — М.: БИНОМ. Лаборатория знаний, 2008. — 655 с.
8. Gorban A. et al. Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007.
9. A. N. Gorban, A. Y. Zinovyev, Principal Graphs and Manifolds, Из: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, Olivas E.S. et al Eds. Information Science Reference, IGI Global: Hershey, PA, USA, 2009. 28-59.
10. Zinovyev A. Vizualizacija mnogomernykh dannyh [Visualization of multidimensional data]. Krasnoyarsk, publ. NGTU. 2000. 180 p. [In Russian]
11. Zinovyev A. Data visualization in political and social sciences, In: SAGE «International Encyclopedia of Political Science», Badie, B., Berg-Schlosser, D., Morlino, L. A. (Eds.), 2011.
12. Питенко А.А. Нейросетевой анализ в геоинформационных системах. Красноярск, Изд. КГТУ, 2000. 97 с.
13. Россиев А.А. Итерационное моделирование неполных данных с помощью многообразий малой размерности. Красноярск, Изд. КГТУ, 2000. 83 с.
14. ViDaExpert, <http://bioinfo.curie.fr/projects/vidaexpert>, last accessed (01 March 2020).
15. Niedoba T., Multi-parameter data visualization by means of principal component analysis (PCA) in qualitative evaluation of various coal types, Physicochemical Problems of Mineral Processing, vol. 50, iss. 2, pp. 575-589, 2014.
16. H. Shaban, S. Tavoularis, Identification of flow regime in vertical upward air–water pipe flow using differential pressure signals and elastic maps, International Journal of Multiphase Flow 61 (2014) 62-72.
17. H. Shaban, S. Tavoularis, Measurement of gas and liquid flow rates in two-phase pipe flows by the application of machine learning techniques to differential pressure signals, International Journal of Multiphase Flow 67(2014), 106-117
18. M. Resta, Computational Intelligence Paradigms in Economic and Financial Decision Making, Series Intelligent Systems Reference Library, Volume 99, Springer International Publishing, Switzerland 2016
19. Bondarev A.E., Bondarenko A.V., Galaktionov V.A., Klyshinsky E.S. Visual analysis of clusters for a multidimensional textual dataset / Scientific Visualization. V.8, № 3, pp.1-24, 2016, URL: <http://sv-journal.org/2016-3/index.php?lang=en>
20. A.E. Bondarev, A.V. Bondarenko, V.A. Galaktionov (2018) Visual analysis procedures for multidimensional data. Scientific Visualization 10.4: 109 - 122, DOI: 10.26583/sv.10.4.09 <http://www.sv-journal.org/2018-4/09?lang=en>
21. Bondarev, A. E.: The procedures of visual analysis for multidimensional data volumes, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W12, 17-21, <https://doi.org/10.5194/isprs-archives-XLII-2-W12-17-2019>, 2019.
22. Bondarev A.E. Visual analysis and processing of clusters structures in multidimensional datasets // Proceedings of the 2nd International ISPRS Workshop on PSBB, 15–17 May 2017, Moscow, Russia, ISPRS Archives, Volume XLII-2/W4, 2017, pp.151-154. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W4/151/2017/>
23. Бондарев А.Е., Галактионов В.А., Шапиро Л.З. Обработка и визуальный анализ многомерных данных / Научная визуализация, Т.9, № 5, 2017, с. 86-104. <http://sv-journal.org/2017-5/08/index.php?lang=ru>

24. Alexander Bondarev, Alexander Bondarenko, Vladimir Galaktionov, Lev Shapiro. Visual Analysis of Textual Information on the Frequencies of Joint Use of Nouns and Adjectives // CEUR Workshop Proceedings, V. 2744, Proc. of the 30th International Conference on Computer Graphics and Machine Vision GraphiCon 2020, Saint Petersburg, Russia, September 22-25, 2020, p. paper20-1 — paper20-10, DOI: 10.51130/graphicon-2020-2-3-20
25. A.E. Bondarev, A.V. Bondarenko, V.A. Galaktionov. Visual Analysis of Text Data Volume by Frequencies of Joint Use of Nouns and Adjectives (2020). Scientific Visualization 12.4: 9 - 22, DOI: 10.26583/sv.12.4.02
26. Bondarev, A. E., Bondarenko, A. V., and Galaktionov, V. A.: Visual analysis of text data collections by frequencies of joint use of words, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIV-2/W1-2021, 21–26, 2021. <https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-21-2021>

Visualization of Points of a Multidimensional Information Text Array on an Elastic Map for Assessing the Cluster Structure of Data

A.E. Bondarev^{1,A}

Keldysh Institute of Applied Mathematics RAS

¹ ORCID: 0000-0003-3681-5212, bond@keldysh.ru

Abstract

The article presents the results of computational experiments on displaying the points of the original multidimensional information array on the elastic map scan to assess the relative positions of semantic proximity areas in order to improve the processing of text information. Elastic maps are considered as a tool for providing analytical work with text information. As previous works show, in order to obtain the required distances corresponding to the cluster picture of the studied multidimensional volume, it is necessary to use the distances on the elastic map, which reflects the cluster portrait of the studied multidimensional data volume. The paper presents the cluster structures of points of the studied multidimensional volume obtained in this way on the elastic map scan in the plane of the first two principal components. An analysis of the relative positions of clusters of different configurations at different points in time is presented.

Keywords: Multidimensional text data, cluster structure, elastic maps, cluster position analysis.

References

1. Thomas J. and Cook K. 2005 Illuminating the Path: Research and Development Agenda for Visual Analytics (IEEE-Press)
2. Wong PC, Thomas J. Visual Analytics // IEEE Computer Graphics and Applications. 2004. V. 24, N. 5. - P. 20-21.
3. Keim D., Kohlhammer J., Ellis G. and Mansmann F. (Eds.) Mastering the Information Age – Solving Problems with Visual Analytics, Eurographics Association, 2010.
4. Kielman, J. and Thomas, J. (Guest Eds.) (2009). Special Issue: Foundations and Frontiers of Visual Analytics / Information Visualization, Volume 8, Number 4, p. 239-314.
5. T. Kohonen, Self-Organizing Maps (Third Extended Edition), New York, 2001, 501 pages.
6. Debock G., Kohonen T. Analysis of financial data using self-organizing maps, Alpina Publisher, 2001, 317 p.
7. Kohonen T. Self-organizing maps. - M.: BINOM. Laboratory of knowledge, 2008. - 655 p.
8. Gorban A. et al. Principal Manifolds for Data Visualization and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007.
9. AN Gorban, AY Zinovyev, Principal Graphs and Manifolds, From: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, Olivas ES et al Eds. Information Science Reference, IGI Global: Hershey, PA, USA, 2009. 28-59.
10. Zinovyev A. Vizualizacija multidimensional dannyh [Visualization of multidimensional data]. Krasnoyarsk, publ. N.G.T.U. 2000. 180 p. [In Russian]

11. Zinovyev A. Data visualization in political and social sciences, In: SAGE “International Encyclopedia of Political Science”, Badie, B., Berg-Schlosser, D., Morlino, LA (Eds.), 2011.
12. Pitenco A.A. Neural network analysis in geoinformation systems. Krasnoyarsk, Publ. KSTU, 2000. 97 p.
13. Rossiev A.A. Iterative modeling of incomplete data using low-dimensional manifolds. Krasnoyarsk, KSTU Publ., 2000. 83 s.
14. ViDaExpert, <http://bioinfo.curie.fr/projects/vidaexpert>, last accessed (01 March 2020).
15. Niedoba T., Multi-parameter data visualization by means of principal component analysis (PCA) in qualitative evaluation of various coal types, Physicochemical Problems of Mineral Processing, vol. 50, iss. 2, pp. 575-589, 2014.
16. H. Shaban, S. Tavoularis, Identification of flow regime in vertical upward air–water pipe flow using differential pressure signals and elastic maps, International Journal of Multiphase Flow 61 (2014) 62-72.
17. H. Shaban, S. Tavoularis , Measurement of gas and liquid flow rates in two-phase pipe flows by the application of machine learning techniques to differential pressure signals, International Journal of Multiphase Flow 67(2014), 106-117
18. M. Resta, Computational Intelligence Paradigms in Economic and Financial Decision Making, Series Intelligent Systems Reference Library, Volume 99, Springer International Publishing, Switzerland 2016
19. Bondarev AE, Bondarenko AV, Galaktionov VA, Klyshinsky ES Visual analysis of clusters for a multidimensional textual dataset / Scientific Visualization. V.8, No. 3, pp.1-24, 2016, URL: <http://sv-journal.org/2016-3/index.php?lang=en>
20. AE Bondarev, AV Bondarenko, VA Galaktionov (2018) Visual analysis procedures for multidimensional data. Scientific Visualization 10.4: 109 - 122, DOI: 10.26583/ sv.10.4.09 <http://www.sv-journal.org/2018-4/09?lang=en>
21. Bondarev, A.E.: The procedures of visual analysis for multidimensional data volumes, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W12, 17-21, <https://doi.org/10.5194/isprs-archives-XLII-2-W12-17-2019>, 2019.
22. Bondarev AE Visual analysis and processing of clusters structures in multidimensional datasets // Proceedings of the 2nd International ISPRS Workshop on PSBB, 15–17 May 2017, Moscow, Russia, ISPRS Archives, Volume XLII-2/W4, 2017, pp.151-154. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W4/151/2017/>
23. Bondarev A.E., Galaktionov V.A., Shapiro L.Z. Processing and visual analysis of multidimensional data / Scientific visualization, Vol.9, No.5, 2017, pp. 86-104. <http://sv-journal.org/2017-5/08/index.php?lang=ru>
24. Alexander Bondarev, Alexander Bondarenko, Vladimir Galaktionov , Lev Shapiro. Visual Analysis of Textual Information on the Frequencies of Joint Use of Nouns and Adjectives // CEUR Workshop Proceedings, V. 2744, Proc. of the 30th International Conference on Computer Graphics and Machine Vision GraphiCon 2020, Saint Petersburg, Russia, September 22-25, 2020, p. paper20-1 - paper20-10, DOI: 10.51130/graphicon-2020-2-3-20
25. AE Bondarev, AV Bondarenko, VA Galaktionov. Visual Analysis of Text Data Volume by Frequencies of Joint Use of Nouns and Adjectives (2020). Scientific Visualization 12.4: 9 - 22, DOI: 10.26583/sv.12.4.02
26. Bondarev, AE, Bondarenko, AV, and Galaktionov, VA: Visual analysis of text data collections by frequencies of joint use of words, Int. Arch. Photogramm . Remote Sens. Spatial Inf. Sci., XLIV-2/W1-2021, 21–26, 2021. <https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-21-2021>