

Процедуры визуального анализа многомерных данных

А.Е. Бондарев¹, А.В. Бондаренко^{2*}, В.А. Галактионов³

Институт прикладной математики им. М.В. Келдыша РАН
* ФГУП ГОСНИИАС

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

² ORCID: 0000-0003-4765-6034, cod@fgosniias.ru

³ ORCID: 0000-0001-6460-7539, vlgal@gin.keldysh.ru

Аннотация

В работе рассматриваются задачи визуального анализа многомерных наборов данных. Для визуального анализа применяется подход построения упругих карт. В работе приведены результаты применения упругих карт для визуального анализа многомерных наборов данных различного происхождения. Описан ряд процедур обработки данных, позволяющих улучшить полученные результаты – предварительная фильтрация данных, удаление отделившихся кластеров (флотация). Для решения проблемы масштабируемости, когда упругая карта подстраивается, как под область сгущения точек данных, так и к отдельно расположенным точкам облака данных, применяется подход quasi-Zoom. Рассматриваются примеры для многомерных данных различного происхождения – для анализа образцов полезных ископаемых, для анализа частот совместного употребления слов при обработке текстов, для анализа многомерного массива ошибок в задаче сравнения точности солверов программного пакета при моделировании обтекания конуса.

Ключевые слова: многомерные данные, визуальный анализ, упругие карты.

1. Введение

В анализе многомерных данных особое место занимают задачи классификации. При решении задач классификации очень полезными оказываются подходы визуальной аналитики, являющиеся синтезом нескольких алгоритмов понижения размерности и визуального представления многомерных данных во вложенных в исходный объем многообразиях, имеющих меньшую размерность.

К таким алгоритмам можно отнести отображение исходного многомерного объема в упругих картах [2,6,7] с разными свойствами упругости или эластичности. Эти методы позволяют тем или иным образом выделить из исходного многомерного объема данных содержащуюся в нем кластерную структуру. Авторами подхода [2,6,7] разработан программный комплекс ViDaExpert [3],

позволяющий проводить построение и визуальное представление упругих карт. Основные функциональные особенности данного программного комплекса подробно описаны в [2].

Интерес к упругим картам появился у нас в процессе реализации проекта по разработке вычислительной технологии для построения, обработки, анализа и визуального представления многомерных параметрических решений задач газовой динамики. Вычислительная технология реализована как единая технологическая цепочка алгоритмов производства, обработки, визуализации и анализа многомерных данных. Такая технологическая цепочка может рассматриваться как прототип обобщенного вычислительного эксперимента для нестационарных задач вычислительной газовой динамики. В итоге подобный обобщенный вычислительный эксперимент позволит получать решение не од-

ной отдельно взятой задачи, а решение для целого класса задач, задаваемого диапазонами изменения определяющих параметров. Также следует отметить универсальность подобного обобщенного вычислительного эксперимента. Он может быть применен к широкому кругу задач математического моделирования нестационарных процессов. Практическая реализация подобного обобщенного эксперимента может обеспечивать организацию крупномасштабных промышленных расчетов. Описание элементов реализованной вычислительной технологии приведено в работах [4,5].

На практике упругие карты оказались полезным и достаточно универсальным инструментом, что позволило применять их к многомерным объемам данных разного типа. Данный подход был применен к задачам анализа текстовой информации, где в качестве числовых характеристик выступали частоты употребления слов [1], а также к задачам анализа образцов полезных ископаемых [9].

2. Упругие карты

Идеология и алгоритмы реализации построения упругих карт подробно представлены в работах [2,6]. Подобная карта представляет собой систему упругих пружин, вложенную в многомерное пространство данных. Этот подход основывается на аналогии с задачами механики: главное многообразие, проходящее через «середины» данных, может быть представлено как упругая мембрана или пластинка. Метод упругих карт формулируется как оптимизационная задача, предполагающая оптимизацию заданного функционала от взаимного расположения карты и данных.

Согласно [2], основой для построения упругой карты является двумерная прямоугольная сетка G , вложенная в многомерное пространство, которая аппроксимирует данные и обладает регулируемыми свойствами упругости по отношению к растяжению и изгибу.

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. Применение упругих карт позволяет более точно и четко определять кластерную структуру изучаемых многомерных объемов данных.

Следует отметить, что при построении упругих карт в многомерном облаке данных, состоящем из сгущений и отдельных отдаленных точек, возникает проблема масштабируемости. Упругая карта будет пытаться подстроиться под рассматриваемый объем в целом – как к отдаленным точкам, так и к областям сгущения, что, естественно, не может получиться одинаково хорошо. Для того чтобы решить эту проблему и обеспечить четкое представление о данных в области сгущений в работе [1], был предложен подход, названный quasi-Zoom, заключающийся в вырезании области сгущения из рассматриваемого облака многомерных данных и построения для вырезанной области упругой карты заново.

3. Процедуры обработки многомерных данных при построении упругих карт

Рассмотрим пример построения упругих карт для объема многомерных данных, представляющих собой описание характеристик полезных ископаемых, а именно, трех сортов угля из месторождений Польши [8].

Рассматриваются многомерные данные, представляющие собой точки в многомерном пространстве признаков (характеристик образцов угля). Пространство признаков состоит из следующих характеристик образцов угля – плотность, масса, удельная теплота сгорания, зольность, содержание серы, содержание летучих компонент, содержание влаги.

Таким образом, мы имеем набор точек в 7-мерном пространстве, соответствующих различным образцам угля. В наборе данных отображены три сорта угля. Рассматривается визуальный анализ с помощью применения упругих карт и главных компонент с целью изучения кластеризации многомерного облака данных и разделения сортов угля. Здесь и далее построение и визуальное представление упругих карт реализовано с помощью программного комплекса ViDaExpert [3], подробно описанного в [2].

Для исходного объема строится «мягкая» упругая карта, отображаемая в пространстве, образованном первыми тремя главными компонентами. Красные, зеленые и синие точки соответствуют трем типам угля (см. рис. 1). Далее представляем развертку построенной карты (см. рис. 2) на плоскость, образованную двумя первыми главными компонентами.

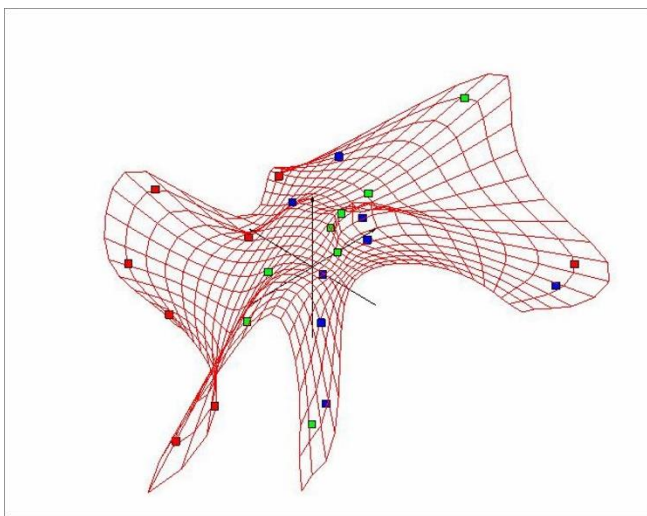


Рис. 1. Построение «мягкой» упругой карты, представляющей три сорта угля.

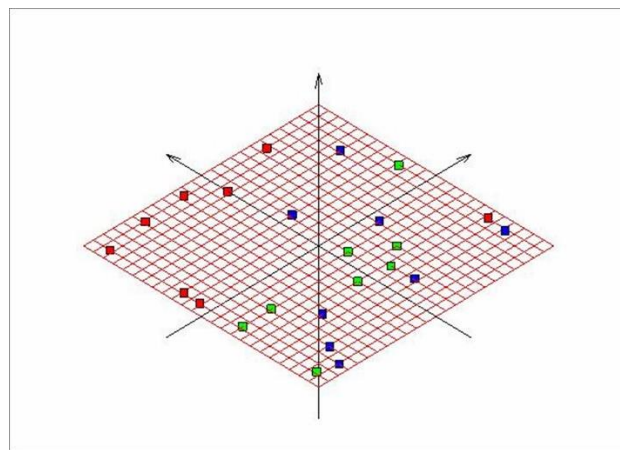


Рис. 2. Развертка «мягкой» упругой карты, представленной на предыдущем рисунке.

По развертке видно, что красные точки образуют отделившийся кластер, за исключением точки в правой части рисунка. Зеленые и синие точки перемешаны.

Для дальнейшего улучшения разделения применим фильтрацию исходного объема данных. Следует отметить, что для некоторых точек в исходном массиве представлены неполные данные, то есть для некоторых образцов информация по ряду характеристик отсутствует или находится в широком диапазоне вариации, а не представлена точно. В частности, данные по размерам образцов представлены неопределенной величиной меньше некоторого или больше некоторого предела (огромные куски или пыль). Попробуем провести фильтрацию данных, то есть убрать все точки, данные по которым представлены нечетким или неполным образом.

Удаление подобных точек из исходного объема приводит к следующим результатам (см. рис. 3).

На рисунке 3 представлена развертка «мягкой» упругой карты для измененного объема данных.

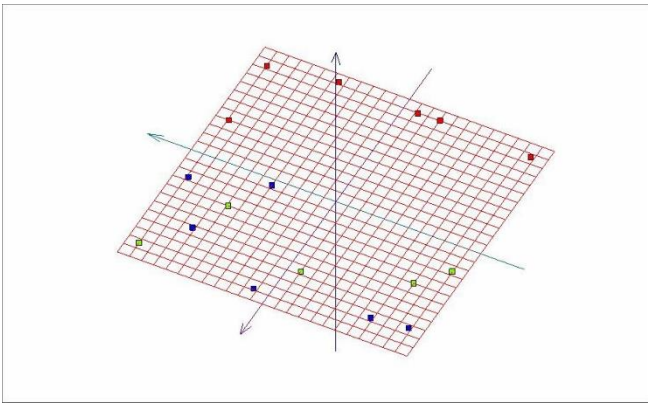


Рис. 3. Развертка «мягкой» упругой карты после фильтрации данных.

Сорт угля, представленный точками красного цвета, полностью отделился после процедуры удаления из исходного объема данных точек с нечетко определенными координатами в 7-мерном пространстве.

Сорта, представленные синими и зелеными точками, остались смешанными. Попробуем еще раз провести ту же процедуру удаления точек из данных. Однако на этот раз исключим из рассматриваемого объема красные точки целиком. Назовем эту процедуру флотацией (от английского термина *flotation*) аналогично термину, применяющемуся при очистке горных пород, когда более легкие фракции всплывают на поверхность и удаляются.

Теперь для четкого разделения двух оставшихся сортов достаточно отобразить точки нового объема данных в пространстве трех первых главных компонент (см. рис. 4).

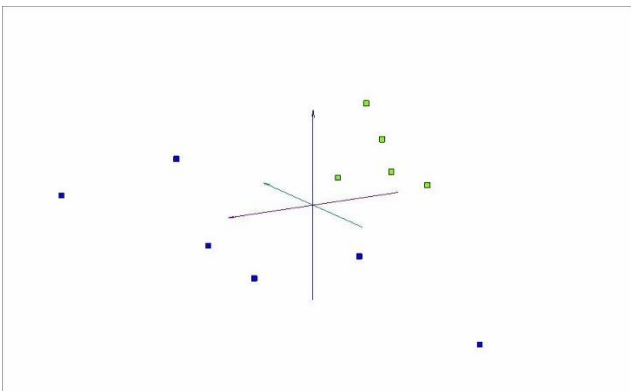


Рис. 4. Оставшиеся точки в пространстве трех первых главных компонент после флотации (удаления отделившегося полностью класса).

Итак, комбинируя построение упругих карт, отображение на вложенные в исходное пространство главные 2D и 3D многообразия и операции удаления нечетких точек и отделившихся классов (фильтрация и флотация данных), можно полностью провести разделение заданных в исходном объеме образцов на три класса, соответствующих трем сортам угля.

Эти же приемы обработки исходных данных послужили основой для разработки приема *quasi-Zoom* при анализе текстовой информации в работе [1]. В этой работе основное внимание уделялось изучению возможности применения методов упругих карт и *t-SNE* (метод стохастического распределения соседей с использованием *t*-распределения Стьюдента) для анализа тематической близости слов русского языка.

Основой предлагаемого метода является анализ непосредственного окружения слов. Основная гипотеза состоит в том, что близкие по смыслу слова должны встречаться в примерно одинаковом контексте. В связи с этим в пространстве признаков они будут находиться на относительно близком расстоянии друг от друга, тогда как отличающиеся слова будут находиться на более удаленном друг от друга расстоянии.

Рассмотрим результаты построения упругих карт для тестового объема [1]. В качестве тестовых данных использовались текстовые корпуса из новостных источников (ленты новостей за определенный период). Процедуры извлечения синтаксически связанных пар слов подробно описаны в [1]. Для первичных тестов было отобрано около 100 глаголов со 155 наиболее связанными с ними существительными. Полученные таким образом данные далее рассматривались как многомерный объем данных, представляющий собой 100 точек в 155-мерном пространстве. Числовые значения получающейся в результате матрицы определяются как частоты совместного употребления. Рассмотрим применение упругих карт к изображению данного объема. На рисунке 5 представлена

развертка «мягкой» упругой карты для данного объема.

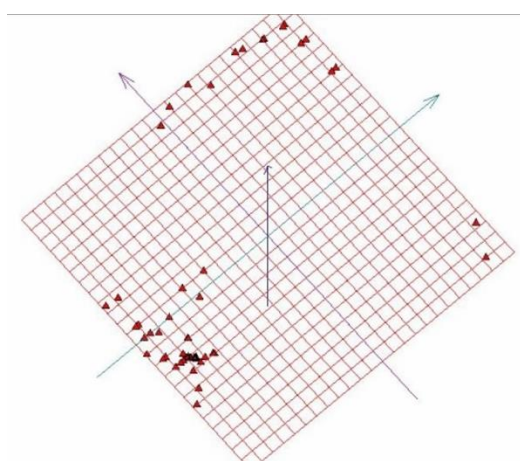


Рис. 5. Построение развертки «мягкой» упругой карты на плоскость первых двух главных компонент [1].

На рисунке видно, что изучаемый объем данных содержит область высокой плотности данных и точки, достаточно далеко отстоящие от этой области.

Именно в таких случаях возникает проблема масштабируемости, описанная ранее. Для решения этой проблемы был разработан подход quasi-Zoom, представленный в работе [1]. При исследовании объема частот совместного употребления глаголов и существительных в работе [1] практическая задача ставилась следующим образом. Нужно было максимально разделить «слипшиеся» точки. Для этой цели был разработан подход quasi-Zoom, который позволил решить эту задачу. Суть этого технологического приема заключается в том, что для более тонкой подстройки необходимо выделять большие кластеры в исследуемом объеме многомерных данных и проводить построение упругих карт для выделенных кластеров отдельно, организуя тем самым эффект, подобный функции «zoom» в современной фототехнике. Это позволит избежать проблем с масштабируемостью, когда упругая карта должна описывать как области сгущения, так и сильно удаленные отдельные точки. Этот прием сродни проиллюстрированным ранее фильтрации и флотации, однако он отлича-

ется от них тем, что имеет четкую последовательность применения.

Для улучшения разделения в области сгущения применим подход quasi-Zoom. Из исходного многомерного объема данных вырежем отделившиеся точки верхней части развертки. К получившемуся в результате этой процедуры новому объему многомерных данных заново применим построение упругой карты. При этом упругую карту будем строить сразу максимально мягкой.

После первого применения подхода quasi-Zoom разделение точек удалось значительно улучшить. Однако под точкой, соответствующей наивысшей плотности данных скрываются еще 37 «слипшихся» слов – глаголов, приведенных на рисунке 6.

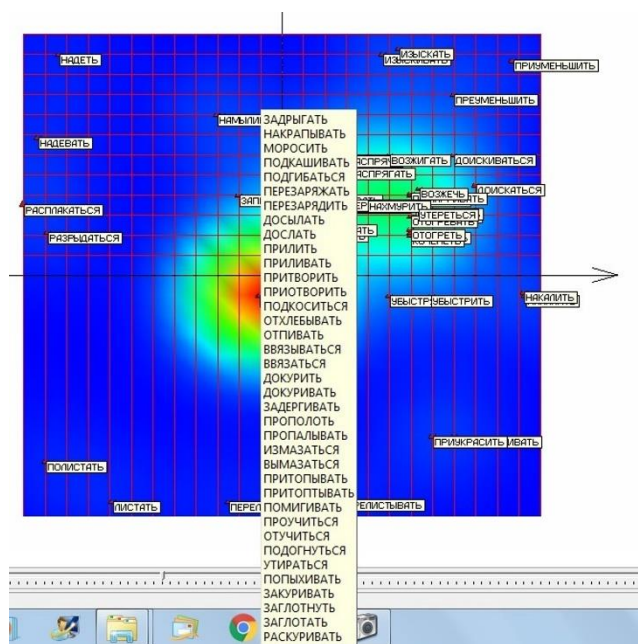


Рис. 6. Список «слипшихся» глаголов в точке наивысшей плотности данных после первого применения подхода quasi-Zoom с раскраской по плотности.

Для того, чтобы обеспечить разделение оставшихся 37 глаголов в точке наивысшей плотности данных, применим прием quasi-Zoom еще раз. Снова вырезаются отделившиеся точки таким образом, чтобы в результирующем многомерном объеме данных остались только 37 «слипшихся точек».

Повторно проводим для получившегося нового многомерного объема данных из 37 точек построение максималь-

но мягкой упругой карты из сетки с таким же количеством узлов, как и на предыдущем этапе. Большую часть точек удалось разделить, однако в зоне наивысшей плотности данных осталось еще 17 «слипшихся» точек.

Проводим процедуру quasi-Zoom в третий раз для оставшихся 17 точек. Результаты представлены на рисунке 7. В результате третьего применения процедуры удалось отделить еще 4 точки. Непосредственный анализ частот встречаемости для оставшихся 13 точек показал, что все они имеют одинаковые координаты по всем измерениям равные нулю. Следовательно, их разделение невозможно в принципе.

Таким образом, процесс разделения точек в тестовом многомерном объеме данных с помощью построения упругих карт и применения процедуры quasi-Zoom завершен полностью и успешно. Удалось разделить все точки многомерного объема данных, имеющие различные и отличающиеся от нуля координаты. Апробация приема quasi-Zoom, проведенная в работе [9] показала эффективность разработанного подхода. При этом результаты показали необходимость проведения фильтрации данных на предварительном этапе с целью удаления точек с полностью нулевыми или полностью одинаковыми координатами. Для этой цели был построен автоматический фильтр в виде программного модуля, позволяющий проводить фильтрацию по указанному признаку. Для полного разделения точек рассматриваемого объема потребовалось только два раза последовательно применить quasi-Zoom. Повторное применение quasi-Zoom уже не оставляет слипшихся точек (рисунок 8). Таким образом, можно сделать вывод, что предварительная очистка данных путем фильтрации существенно ускоряет обработку многомерного массива и улучшает качество обработки.

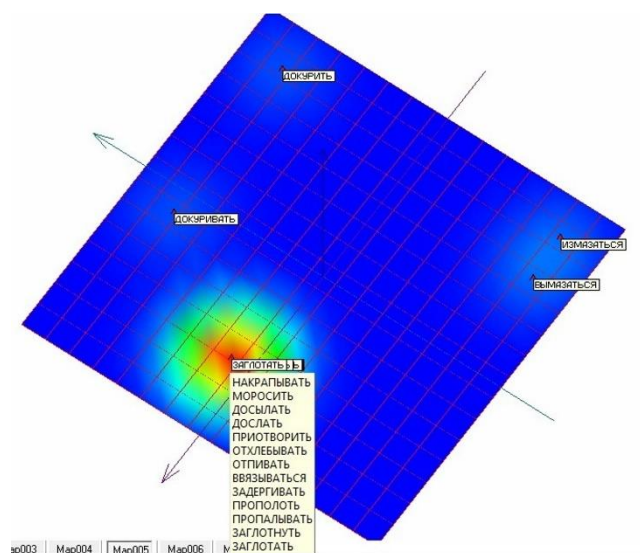


Рис. 7. Развертка упругой карты с раскраской по плотности после третьего применения процедуры quasi-Zoom [1].

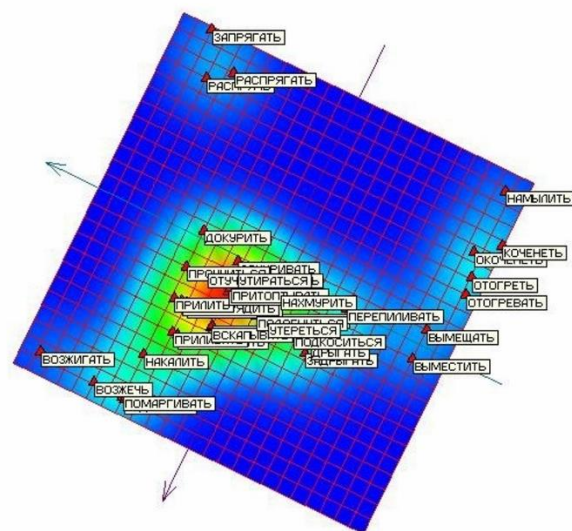


Рис. 8. Развертка упругой карты с раскраской по плотности после фильтрации и двукратного применения процедуры quasi-Zoom [9].

Следующим шагом является повторение вышеописанных процедур для массива большей размерности. Рассмотрим массив, состоящий из 100 точек в пространстве из 353 измерений. Массив строился так же и с теми же самыми глаголами, как это было представлено выше для случая 155 измерений. Но только теперь увеличилось количество существительных, с которыми

рассматривается совместное употребление этих глаголов. Повторим всю процедуру для этого массива. Сначала проведем фильтрацию, а затем построим упругую карту и ее развертку. На рисунках 9, 10 представлены соответственно: набор данных в пространстве главных компонент, развертка упругой карты для этого набора данных.

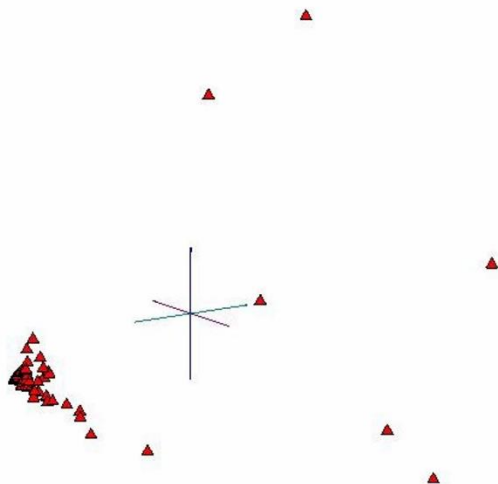


Рис. 9. Набор данных увеличенной размерности в пространстве главных компонент.

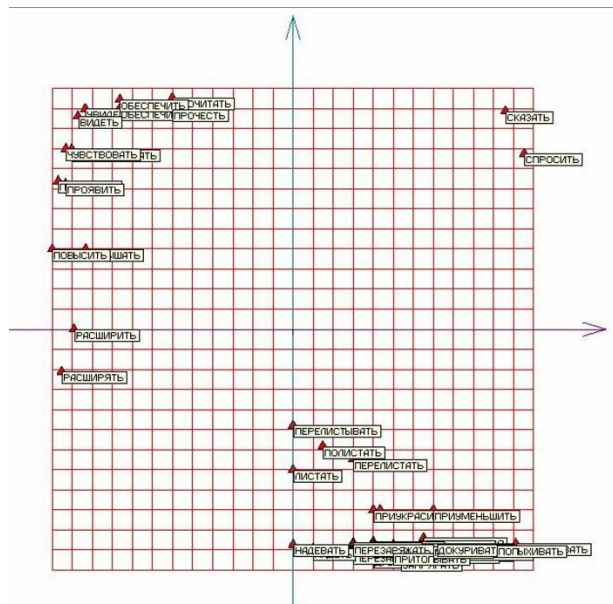


Рис. 10. Развертка упругой карты для набора данных увеличенной размерности.

Следует заметить, что парность схожих глаголов хорошо отражается на развертке упругой карты. Это отражает-

ся на рисунке 11, представляющего часть рисунка 10 крупным планом.

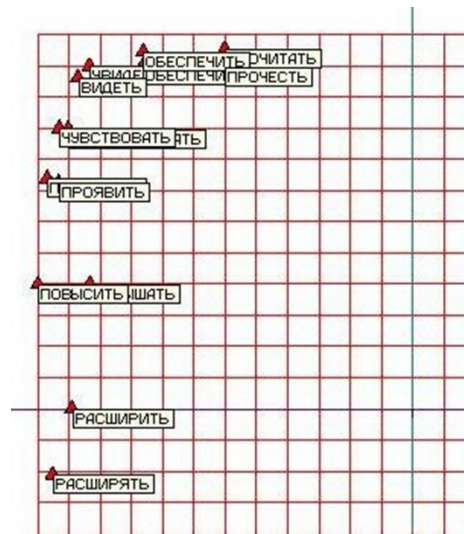


Рис. 11. Крупный план части рисунка 10.

Теперь применим подход quasi-Zoom, удалим отделившиеся точки и построим для оставшихся точек упругую карту, ее развертку на плоскость двух первых главных компонент и раскраску развертки по плотности данных. Результаты построения упругой карты и ее развертки представлены на рисунках 12,13,14.

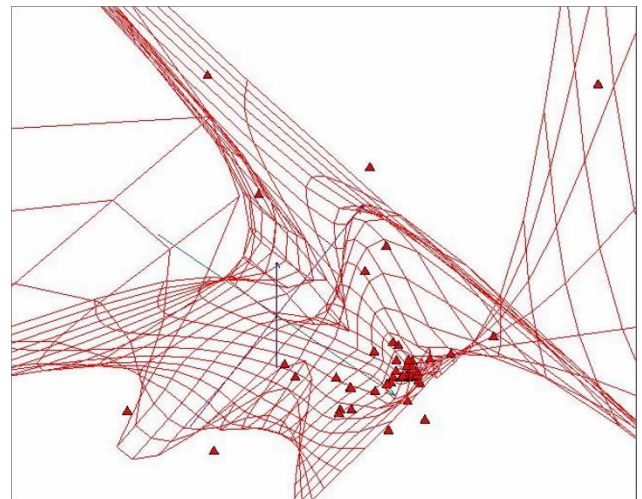


Рис. 12. Упругая карта для данных повышенной размерности после первого применения quasi-Zoom.

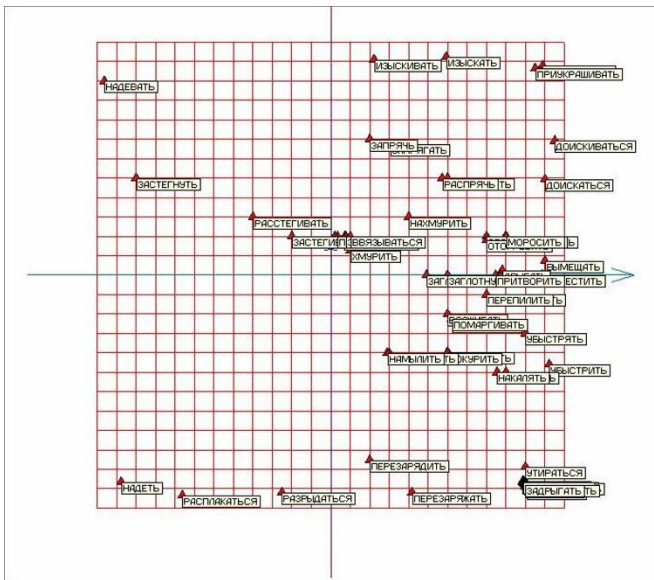


Рис. 13. Развертка упругой карты для данных повышенной размерности после первого применения quasi-Zoom.

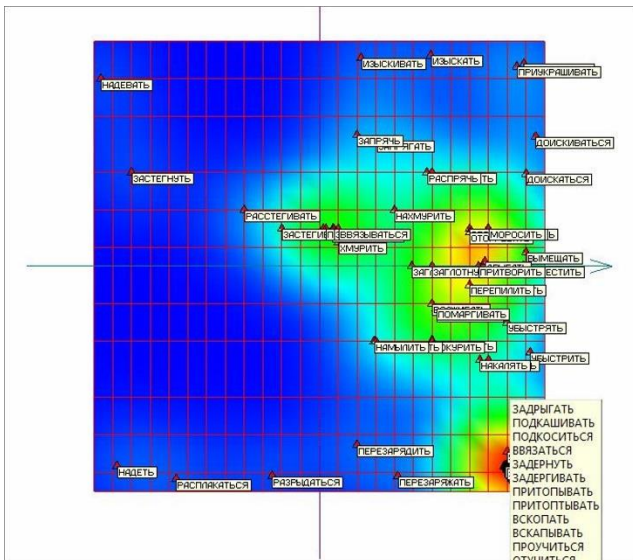


Рис. 14. Развертка упругой карты с раскраской по плотности для данных повышенной размерности после первого применения quasi-Zoom.

Дальнейшее применение подхода позволяет прийти к полному разделению точек изучаемого объема данных. Это может быть проиллюстрировано рисунками 15 и 16, где представлены результаты второго применения процедуры quasi-Zoom.

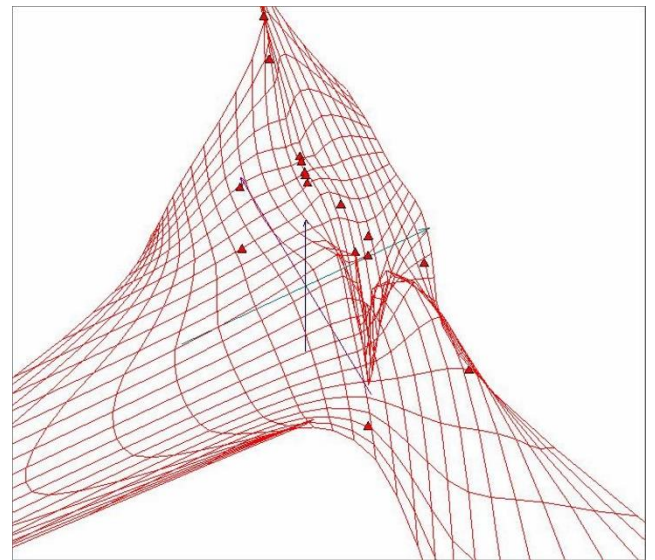


Рис. 15. Упругая карта для данных повышенной размерности после второго применения quasi-Zoom.

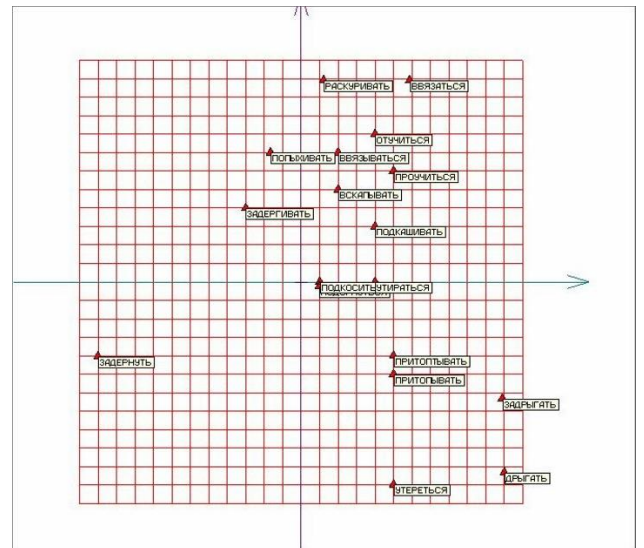


Рис. 16. Развертка упругой карты для данных повышенной размерности после второго применения quasi-Zoom.

Таким образом, применение процедур фильтрации данных и quasi-Zoom позволяет успешно решать задачи разделения «слипшихся» точек многомерном объеме данных.

Следует заметить, что при изучении непосредственно матрицы данных обнаруживается, что лучше всего отделяются точки, имеющие максимальное число ненулевых координат по всем 353 измерениям. В области сгустка оказываются точки, у которых ненулевые измерения составляют одно или два. Остальные равны нулю. Это объяснимо

с точки зрения алгоритма построения главных компонент.

Из полученных результатов можно сделать следующий вывод – увеличение количества измерений с ненулевыми значениями улучшает процесс разделения точек в областях сгущения. Не следует забывать, что предыдущий 155-мерный массив был получен из настоящего рассматриваемого массива путем удаления измерений.

Следующим интересующим нас вопросом являлся вопрос влияния на результат транспонирования рассматриваемого многомерного массива данных. Что будет, если поменять существительные и глаголы местами? Теперь рассмотрим 155 точек, характеризующих существительные, а в качестве измерений у нас будут служить 100 глаголов (87 после фильтрации). Соответственно многомерный массив будет также состоять из частот совместного употребления существительных и глаголов, и мы сможем посмотреть и оценить, насколько схожие в понятийном смысле слова будут близки на развертке упругой карты.

Исходный массив данных представлен на рисунке 17 в пространстве главных компонент. Видно, что при транспонировании принципиально структура данных не поменялась. Все так же мы видим область сгущения и ряд удаленных точек.

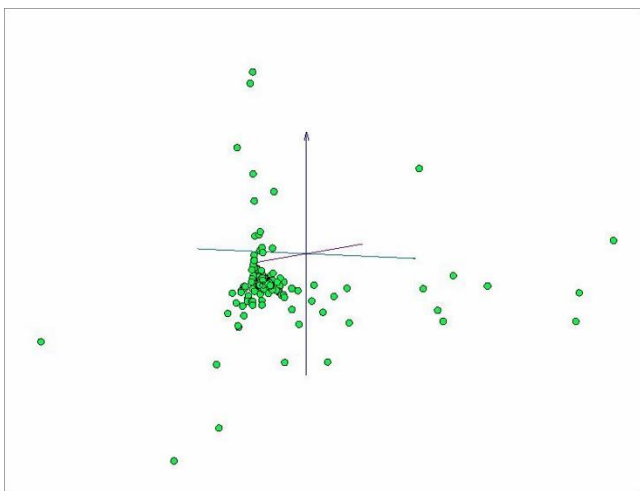


Рис. 17. Транспонированный массив данных в пространстве главных компонент.

На рисунках 18 и 19 представлена построенная для транспонированного массива упругая карта и ее развертка в плоскости первых главных компонент.

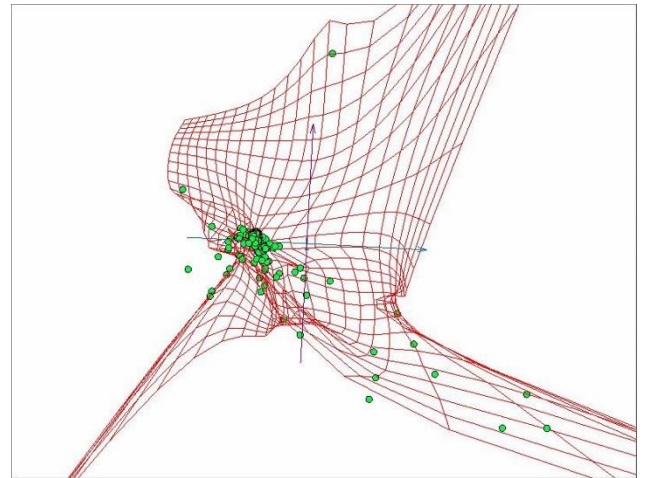


Рис. 18. Упругая карта для транспонированного массива данных.

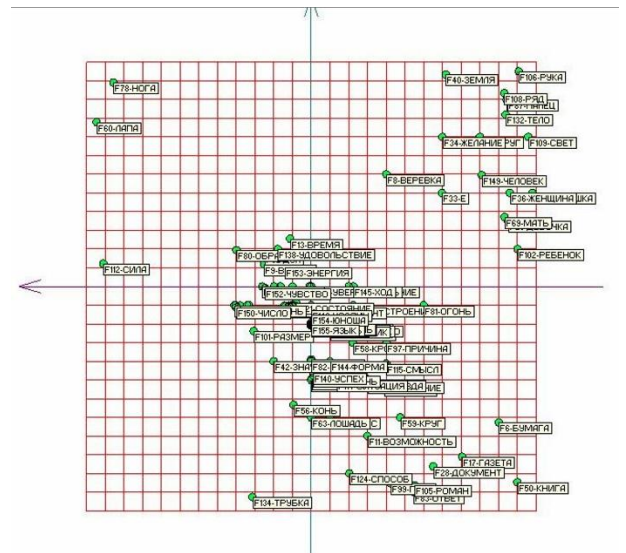


Рис. 19. Развертка упругой карты для транспонированного массива данных.

Здесь в отличие от предыдущих рассмотренных случаев нет парных глаголов, а есть слова, близкие в понятийном смысле. Рассмотрим их крупным планом.

На рисунках 20 и 21 представлены верхняя правая и нижняя правая части развертки упругой карты. На рисунке 20 можно видеть достаточно четко выделенный понятийный кластер – ЧЕЛОВЕК, ЖЕНЩИНА, ДЕВОЧКА, ДЕВУШКА, МАТЬ, РЕБЕНОК. Чуть выше менее четко просматривается смысловой кластер – РУКА, ПАЛЕЦ, ЛИЦО, ТЕЛО.

На рисунке 21 также просматривается кластер понятий – СМЫСЛ, ДОКУМЕНТ, БУМАГА, ГАЗЕТА, КНИГА, РОМАН, ОТВЕТ.

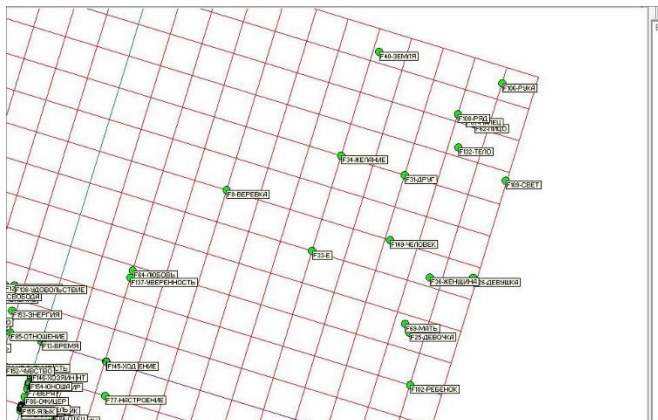


Рис. 20. Верхняя правая часть развертки упругой карты.

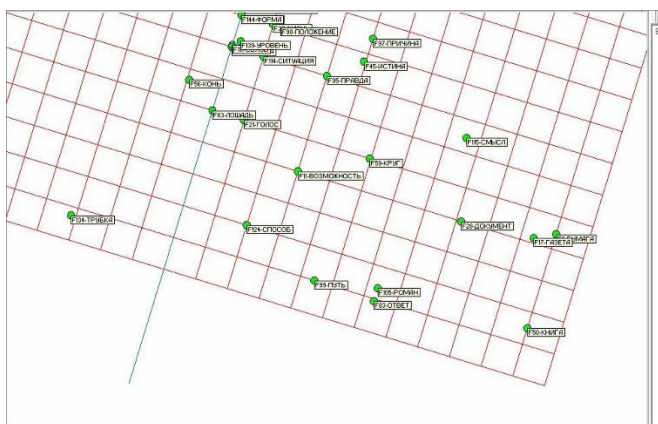


Рис. 21. Нижняя правая часть развертки упругой карты.

Для более тщательного изучения слов в сгущении на центральной части рисунка 19 применим подход quasi-Zoom. На рисунке 22 представлена развертка упругой карты после применения процедуры.

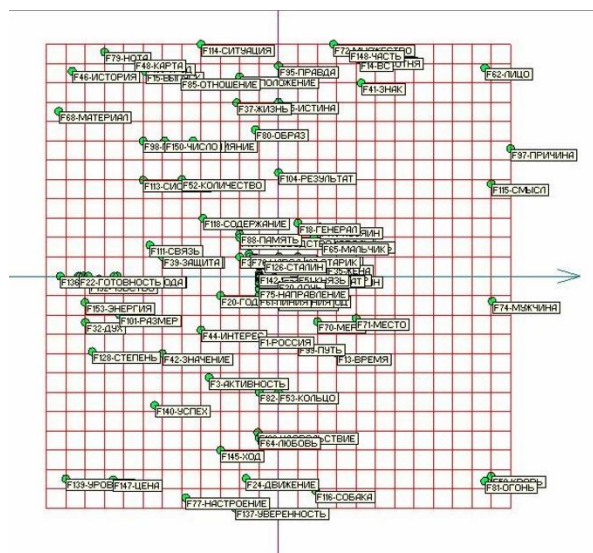


Рис. 22. Развертка упругой карты для транспонированного массива данных после применения quasi-Zoom.

На рисунке 22 можно увидеть ряд потенциальных кластеров из слов, близких в понятийном смысле. Например: ЛЮБОВЬ, УДОВОЛЬСТВИЕ или ПРИЧИНА, СМЫСЛ или УРОВЕНЬ, ЦЕНА. Также можно увидеть интересный понятийный кластер – РОССИЯ, ПУТЬ, ВРЕМЯ, МЕРА, МЕСТО. Следует напомнить, что мы рассматриваем массив, состоящий из частот совместного употребления глаголов и существительных, а совместное употребление отражает, в свою очередь, образовавшиеся устойчивые конструкции слов. При дальнейшем изучении массива в центральной области можно выделить кластер – СОЛДАТ, МУЖ, ОФИЦЕР, КНЯЗЬ, КОМАНДИР, ЮНОША, что показано на рисунке 23.

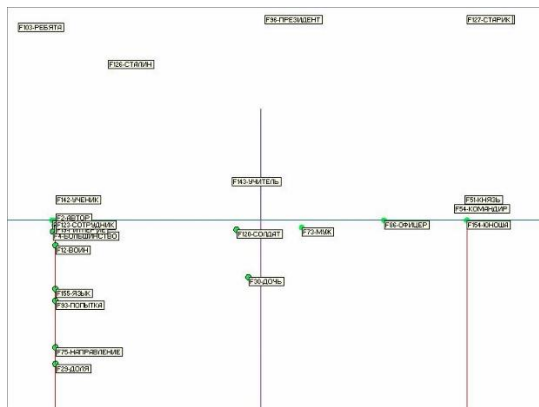


Рис. 23. Центральная часть развертки упругой карты для транспонированного массива данных после применения quasi-Zoom.

Это открывает широкие возможности по анализу и трактовке смысловых группировок для специалистов в этой области. Для более тщательного анализа и трактовки смысловой близости следует развивать данные подходы с точки зрения анализа расстояний между точками многомерного массива для различного задания метрики. Также интересной и важной задачей в этом плане должно стать выяснение влияния выбора метрики.

Для улучшения разделения необходимо в дальнейшем усовершенствовать алгоритм построения упругих карт за счет возможности сгущения исходной сетки карты в областях повышенной плотности данных. Подобная возможность широко используется в задачах математического моделирования сплошных сред.

Таким образом, описанная система процедур позволяет дополнить инструментарий визуальной аналитики набором приемов, позволяющих получить более полное представление об изучаемом массиве многомерных данных. По сути, описанные приемы являются достаточно универсальными и могут применяться к самым разнообразным наборам многомерных данных, независимо от природы их происхождения.

Рассмотрим применение упругих карт к анализу многомерного массива ошибок по сравнению с точным решением. Многомерный массив ошибок получен в параметрических исследованиях, описанных в [10,11]. В этих работах представлены численные результаты сравнения точности работы различных солверов программного пакета OpenFOAM на примере известной задачи невязкого обтекания конуса под нулевым углом атаки. Проводится сравнение результатов, полученных с помощью различных солверов OpenFOAM, с известным численным решением данной задачи при вариации числа Маха набегающего потока и угла полураствора конуса. Приведены результаты расчетов, проведенных для ряда расчетных сеток. В сравнении участвовали 4 солвера программного пакета OpenFOAM –

rhoCentralFoam, pisoCentralFoam, sonicFoam, rhoPimpleFoam. Все эти солверы обладают разными аппроксимационными и вычислительными свойствами. В качестве точного решения для сравнения использовались табличные данные, сравнение проводилось с помощью аналога нормы L_2 .

На рисунке 24 представлен многомерный массив данных для давления, полученных в результате параметрических расчетов, в пространстве первых главных компонент. Желтыми кружками представлены результаты для солвера rhoCentralFoam, красными – для pisoCentralFoam, зелеными – для sonicFoam и голубыми – для rhoPimpleFoam. На рисунке 25 представлена упругая карта для этого же массива, а на рисунке 26 представлена развертка упругой карты.

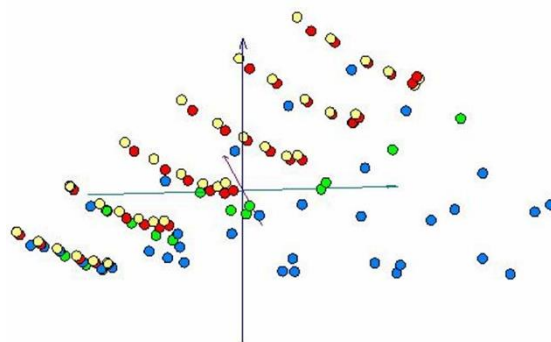


Рис. 24. Массив ошибок по давлению для 4-х солверов OpenFOAM при вариации числа Маха и угла полураствора конуса в пространстве главных компонент.

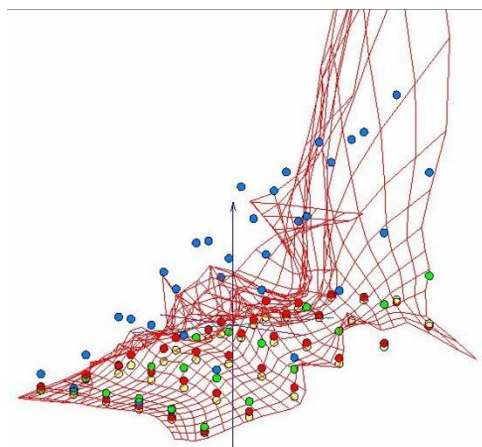


Рис. 25. Упругая карта для массива ошибок по давлению для 4-х солверов OpenFOAM

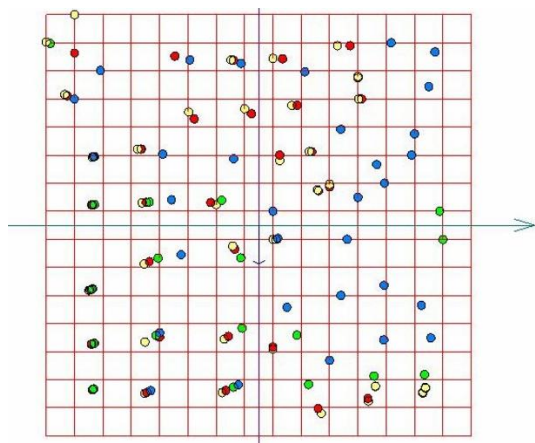


Рис. 26. Развертка упругой карты для массива ошибок по давлению для 4-х солверов OpenFOAM

Из приведенных рисунков можно увидеть, что ошибки для `rhoCentralFoam` и для `pisoCentralFoam` можно грубо аппроксимировать плоскостью, отражающей зависимость ошибки от числа Маха и угла полураствора. Результаты для `sonicFoam` и в особенности для `rhoPimpleFoam` существенно отделяются от результатов для двух первых солверов. Для более точного изучения массива ошибок для солверов в аналоге нормы L_2 предполагается увеличить размерность рассматриваемого массива путем проведения всех сделанных ранее расчетов при вариации угла атаки набегающего потока для данной задачи.

Заключение

Для анализа кластерных структур в многомерных объемах данных использованы технологии построения упругих карт, представляющие собой методы отображения точек исходного многомерного пространства на вложенные многообразия меньшей размерности. Рассмотрен ряд приемов обработки данных, позволяющих улучшить результаты – предварительная фильтрация данных, удаление отделившихся кластеров (флотация). Для решения проблемы масштабируемости, когда упругая карта подстраивается как под область сгущения точек данных, так и к отдельно расположенным точкам облака данных, применяется подход `quasi-Zoom`. Суть подхода заключается в том,

что для более тонкой подстройки в многомерном облаке данных выделяются большие кластеры, после чего проводится отдельное построение упругих карт для выделенных кластеров. Приведены примеры построения упругих карт для многомерных данных различного происхождения – для анализа образцов угля, для анализа частот совместного употребления слов при обработке текстов, для анализа многомерного массива ошибок в задаче сравнения точности солверов программного пакета при моделировании обтекания конуса.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект № 18-11-00215).

Литература

- [1] 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, с.1-24, 2016, URL: <http://sv-journal.org/2016-3/index.php?lang=en>
- [2] Зиновьев А.Ю. Визуализация многомерных данных - Красноярск: Изд-во КГТУ, 2000 - 168 с.
- [3] Программный пакет ViDaExpert <http://bioinfo.curie.fr/projects/vidaexpert> / (дата обращения 01.02.2018).
- [4] Bondarev A.E., Galaktionov V.A. Analysis of Space-Time Structures Appearance for Non-Stationary CFD Problems / *Proceedings of 15-th International Conference On Computational Science ICCS 2015* Reykjavik, Iceland, June 01-03 2015, *Procedia Computer Science*. Vol. 51. P. 1801–1810.
- [5] Bondarev A.E., Galaktionov V.A. Multidimensional data analysis and visualization for time-dependent CFD problems / *Programming and Computer Software*. 2015. Vol. 41. №. 5. P. 247–252. DOI: 10.1134/S0361768815050023
- [6] Gorban A., Kegl B., Wunsch D., Zinovyev A. (Eds.), *Principal Manifolds for Data Visualisation and Dimension Reduction*, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007.

[7] Gorban A.N., Zinovyev A. Principal manifolds and graphs in practice: from molecular biology to dynamical systems / International Journal of Neural Systems. 2010. Vol. 20. №. 3. P. 219–232. DOI: 10.1142/S0129065710002383

[8] 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. 2014. Vol. 50. № 2. P. 575-589.

[9] Bondarev A.E., Galaktionov V.A., Shapiro L.Z. Processing and visual analysis of multidimensional data / Scientific Visualization. V.9, № 5, c.86-104, 2017, DOI: <http://doi.org/10.26583/sv.9.5.08>

[10] Bondarev A.E., Kuvshinnikov A.E. Comparative study of the accuracy for OpenFOAM solvers // Proceedings of Ivannikov ISPRAS Open Conference (ISPRAS), 2017, IEEE, IEEE Xplore, p.132-136 DOI: 10.1109/ISPRAS.2017.00028

[11] Alexander E. Bondarev and Artem E. Kuvshinnikov. Analysis of the Accuracy of OpenFOAM Solvers for the Problem of Supersonic Flow Around a Cone // ICCS 2018, Lecture Notes in Computer Science (LNCS) 10862, pp. 221–230, 2018. https://doi.org/10.1007/978-3-319-93713-7_18