Правительство Российской Федерации

Государственное образовательное бюджетное учреждение высшего профессионального образования

«Национальный исследовательский университет Высшая школа экономики»

Факультет Социологии

Кафедра Методов сбора и анализа социологической информации

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему «Интерпретация результатов типологического анализа на основе фрактального подхода»

Студент группы № 732
Лаврова Екатерина Олеговна

Научный руководитель
Д.т.н., проф. Буховец Алексей Георгиевич

Москва 2010

Содержание

Введение

Актуальность темы. Как известно, задача выделения классов объектов, схожих по каким-либо характерным признакам, является одной из центральных в процессе анализа социологической информации. Однако понятие классификации включает в себя не только разделение объектов на классы. Главная цель социолога при анализе данных – это получение нового знания, такого. «Назначение всякой классификации … заключается, прежде всего, в том, чтобы быть средством лучшего познания изучаемых объектов, о которых ещё не имелось сформировавшихся понятий» [Формальная логика, 1977, с.140].

В таком понимании проблемы классификации встает два вопроса: о лучшей интерпретации классов объектов и о лучшем понимании через классификацию самой структуры данных. Второй вопрос «приводит к тому, что от классификации требуется обязательное раскрытие сущности изучаемого явления на уровне некоторого закона» [Буховец, Математическое моделирование структур многомерных данных в классификационных задачах, 2006].

Первый же вопрос – о лучшей интерпретации классов объектов – требует более практического взгляда на методы, с помощью которых описываются выделенные классы. Для этого обычно применяют традиционный методный арсенал – дескриптивные характеристики объектов, дисперсионный, факторный, регрессионный анализ. При регрессионном анализе, к примеру, к каждому классу аппроксимируется некая функция, обычно линейная. То есть, мы наблюдаем процесс приближения объектов – точек пространства – и линейной функции, которая, по сути, есть часть плоскости.

Говоря о выявлении структуры данных при классификации, нельзя также не упомянуть о принятии априорного предположения о нормальности распределения показателей, которое все чаще подвергается критике. Говоря о распределении данных в пространстве признаков, исследователи все чаще приходят к выводу о том, что нормальное распределение характерно для физических величин, в то время как для экономических и социологических данных характерно распределение гиперболическое. Такой подход хорошо иллюстрируется цитатой: «Природа, говорим мы, в основном, гауссова, социум (человек) негауссов» [Хайтун, 1989, с. 111]. Указанное распределение имеет ряд привлекательных для социологов свойств, таких как более тяжелые хвосты (данные о доходах российских граждан, к примеру, имеют более тяжелые хвосты и этим самым не «вписываются» в законы о нормальном распределении) и возможность учитывать асимметричность данных (на это свойство гиперболического распределения в применении к социальным процессам указывают в первую очередь экономисты [Истигечева, 2006, с. 11]).

Отдельного внимания заслуживает также принятие закона больших чисел, смысл которого заключается в том, что совокупное действие независимых внешних факторов приводит к результату, почти не зависящему от случая. Помимо очевидного удобства закона (чем больше мы сделаем наблюдений и замеров, тем с большей вероятностью в итоге нам удастся избежать ошибок), у него есть некоторая особенность, а конкретнее, условия применения закона больших чисел, а если быть более точным – центральной предельной теоремы, которые «по существу сводятся к требованию, чтобы влияние на сумму отдельных слагаемых было равномерно малым, т.е. чтобы в состав суммы не входили члены, явно преобладающие над совокупностью остальных по своему влиянию на рассеивание суммы. […] Неучтенные случайные факторы, создающие рассеяние, обычно характерны своей равномерностью и отсутствием среди них резко преобладающих» [Вентцель, 1964, с. 287].

Очевидно, что в социальных процессах часто встречаются факторы, явно преобладающие (например, центральный банк на рынке валют). При снятии некоторых ограничений в законе больших чисел появляется логнормальное распределение. Однако и здесь подробно не рассматривается факт большего воздействия одних переменных на объект и меньшего – других. Если же принять положение о наличии доминирующего фактора (т.е. фактора, величина которого больше суммы остальных), то в математических моделях это приводит к тому, что нарушается неравенство треугольника, и это, в свою очередь, влечет за собой отказ от изотропного метрического пространства и рассмотрению структуры объектов как некоторой фрактальной структуры [Буховец, 2010].

Рассмотренные выше доводы делают актуальной проблему построения содержательной интерпретации групп объектов, получаемых в ходе их классификации, а также нахождения лучшей структуры данных. В ходе данной работы автором предлагается использовать в данных целях фрактальный подход. Интерпретация классов в рамках этого подхода осуществляется путем аппроксимирования классов не функцией плотности или гиперплоскостью в признаковом пространстве, а подобным специальным образом сгенерированным фрактальным множеством.

Целью данной работы является исследование возможностей использования фрактального подхода при анализе эмпирических данных в задачах классификации.

Для реализации цели работы необходимо решение следующих **задач:**

1. Изучение моделей построения фрактальных структур в признаковых пространствах классификации, рассмотрение структур социологических данных с точки зрения фрактального подхода.
2. Исследование особенностей применения кластерного анализа в социологических исследованиях.
3. Разработка методики подхода, позволяющего использовать фрактальные структуры для получения и интерпретации результатов классификации.

4. Проверка предложенных рекомендаций на реальных эмпирических данных

Теоретическим объектом работы является фрактальный подход, **предметом** – применение фрактального подхода в задачах интерпретации результатов типологического анализа.

Глава 1. Методологические аспекты применения фрактального подхода в задачах типологического анализа.

1.1. Определение фрактала

"Мендель никогда не открыл бы законов генетики, даже располагая самыми современными статистическими методами и данными, если бы он не знал, что ищет."
Ф.Мостеллер

Перед рассмотрением структуры данных социальных процессов как фрактальной структуры, необходимо прежде всего дать определение фракталу. Фрактал (от лат. fractal – дробный, сломанный) — термин, введенный французским математиком Бенуа Мандельбротом, основателем и ведущим исследователем в области фрактальной геометрии, в 1975 г. Он обозначает бесконечно самоподобную геометрическую фигуру, каждый фрагмент которой повторяется при уменьшении масштаба. Основными свойствами фракталов являются:

- Масштабная инвариантность: форма объектов не меняется от точки нахождения наблюдателя (в отличие, к примеру, от окружности, превращающейся в линию при ближайшем рассмотрении);
- Самоподобие или самоафинность: малая часть объекта есть «уменьшенная копия» большой ее части, и при увеличении может быть наложена на эту большую часть или на объект в целом;
- Фракталы обладают дробной метрической размерностью или метрической размерностью, превосходящей топологическую.

Топологическая размерность фигуры – это количество координат, необходимых для определения положения лежащей на этой фигуре точки. То есть, у линии топологическая размерность равна 1, у плоской фигуры – 2, у объемной фигуры – 3. Метрическая же размерность зависит от единиц измерения. Например, рассмотрим задачу измерить длину извилистой береговой линии. Если измерить ее километрами, получится одно значение. Но при приближении (увеличении масштаба) увеличивается ее длина, так как обнаруживаются более мелкие извилины, и при измерении ее с применением более мелких единиц длина береговой линии окажется больше (к примеру, длина такой линии в метрах будет равна 145 м, а в миллиметрах – 145355 мм). Такая линия является чем-то промежуточным между математической линией и математической плоскостью и имеет размерность $1<D<2$. Для вычисления размерности фракталов используется размерность Хаусдорфа[1] для ограниченного множества и Хаусдорфа-Безиковича – для неограниченного. На практике такое измерение сводится чаще всего к измерению клеточной размерности, когда объект покрывается клетками с единичной стороной.

Следует иметь в виду, что в дальнейшем сам Мандельброт отказался от этого определения (включающее в себя обязательность дробной метрической размерности), т.к. были получены фрактальные структуры с целочисленной размерностью [26]. В конечном счете, определение фрактала, по Мандельброту, сводится к тому, что фрактал – это объект, самоподобный или самоафинный в том или ином смысле.

Построение самоподобного объекта можно проиллюстрировать на примере фрактала, известного как «треугольник Серпинского» (рис.1). Равносторонний треугольник M0 делится прямыми, параллельными его сторонам, на 4 равных равносторонних треугольника. Из треугольника удаляется центральный треугольник. Получается множество M1, состоящее из 3 оставшихся треугольников "первого ранга". Поступая точно так же с

[1] В рамках данной работы не рассматривается

каждым из треугольников первого ранга, получим множество М2, состоящее из 9 равносторонних треугольников второго ранга. Продолжая этот процесс бесконечно, получим бесконечную последовательность множеств

$$M_0 \supset M_1 \supset \ldots \supset M_n \supset \ldots,$$

пересечение членов последовательности которой есть треугольник Серпинского.

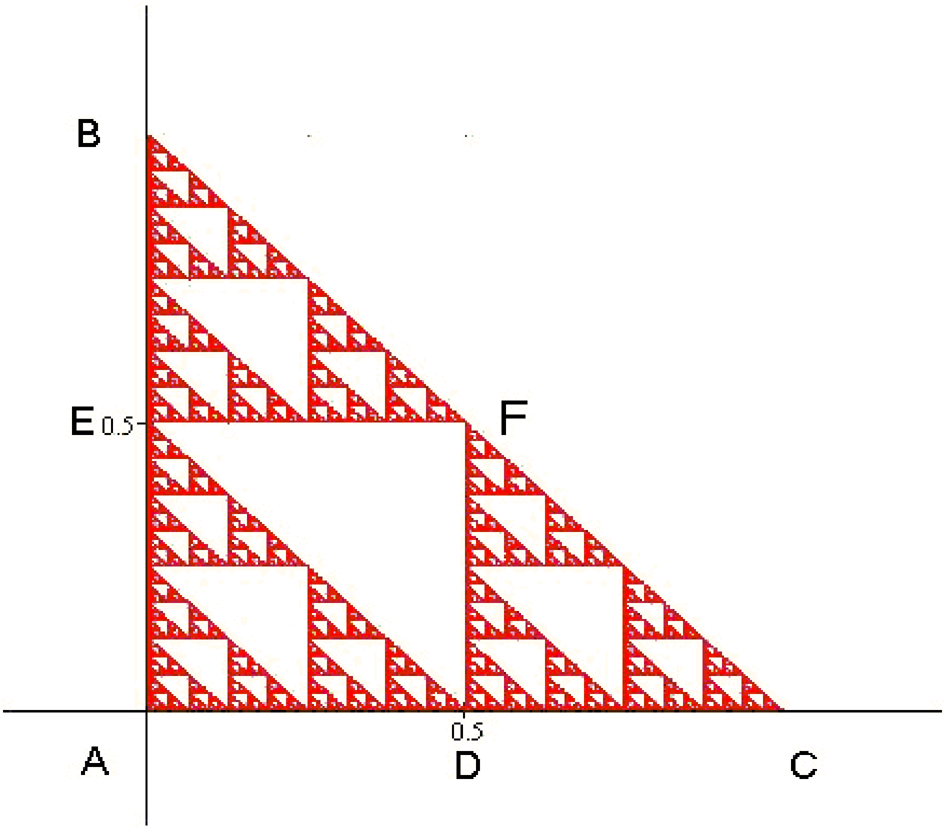


Рис. 1. Построение фрактального множества.

Наряду с этим рассмотрим следующую задачу. Отрезок [0; 1] случайным образом делится на три части. Какова вероятность, что из полученных частей можно составить треугольник? Для решения этой задачи введем в рассмотрение положительные величины: x, y и z, обозначающие длины отрезков, полученных при делении. В силу выполнения очевидного условия x + y + z = 1 в дальнейшем оставим только две независимые величины x и y. Множество допустимых значений задачи будет представлено следующей системой неравенств $x > 0$; $y > 0$; $x + y < 1$ (1), которая будет соответствовать ранее рассмотренному треугольнику ABC. Очевидно, что можно установить взаимнооднозначное соответствие между способами деления единичного отрезка и точками построенного треугольника. Учитывая требование неравенства треугольника можно составить систему неравенств $x < 1/2$; $y < 1/2$; $x + y > 1/2$. Решение этой системы будет представлено в виде DEF, являющегося центральной частью исходного ABC. Отсюда следует вывод, что вероятность того, что из полученных после деления отрезков можно будет образовать треугольник, равняется 0,25.

Здесь, однако, интересен другой момент: в то время как точки DEF, соответствуют делению отрезка, позволяющему составить треугольник, точки других оставшихся треугольников соответствуют делениям, для которых условия неравенств системы (1) не выполняются. Другими словами, DEF образуют точки, деление которыми позволяет получить отрезки, каждый из которых «не будет явно преобладать над совокупностью остальных», а в отношении трех остальных треугольников это утверждение не имеет места. Иначе говоря, в каждом из оставшихся треугольниках существует одна сторона, преобладающая над суммой остальных.

Если теперь продолжить этот процесс, взяв каждый из больших отрезков и разделить его опять на три части, а затем определить возможность получения треугольника, то эта операция вновь сведется к удалению из каждого оставшегося треугольника центральной его части. В результате продолжения таких действий мы опять будем получать множества отрезков, в которых будет иметься отрезок, доминирующий над всеми остальными. А это в свою очередь, заметим, не позволит составить из таких отрезков треугольник, т.е. полученные отрезки не будут «равномерно малыми». Очевидно, что организованная таким образом процедура в пределе сведется к описанной выше процедуре построения треугольника Серпинского. Но при этом появляется возможность интерпретировать остающиеся после выполнения каждого шага построения множества точек как множества, имеющие в своем составе доминирующий элемент – в данном случае отрезок, длина которого превосходит сумму остальных.

Таким образом, в структуре фрактала мы можем наблюдать доминирующий фактор, которого не было при условии выполнения неравенства треугольника.

1.2. Примеры фрактальных структур в данных результатов социальных процессов.

1.2.1. Примеры фрактальных структур

Один из самых ярких и понятных социологу примеров фракталов в социологии – репрезентативная выборка в опросах общественного мнения, где б*о*льшая генеральная совокупность изучается с помощью исследования меньшей выборочной совокупности.

Фрактальные образования можно наблюдать на уровне взаимодействия, взаимоотношения элементов, составляющих социальный механизм. Сущностной основой такого понимания лежит закономерность взаимозависимости отношений и общества и личности следующего вида: человек – микрокосм истории общества. Понятно, что в самом общем случае человек является микрокосмом Вселенной, частью которой выступает общество в его динамике. Данная закономерность наблюдается во фрактальном осмыслении явлений окружающего нас мира [Волков, Мостовая. Социология. М.: 1998].

В качестве примеров фрактальных образований в данных социальных измерений можно рассмотреть ранговые распределения, которые можно встретить, например, при изучении доходов населении, продуктивности научных работников, распределения городов страны по численности их населения, и проч. Такого рода распределения хорошо аппроксимируются гиперболическим законом распределения.

Примеры самоподобных структур (масштабно-инвариантных построений) можно легко обнаружить в организационных структурах иерархических групп людей. Признаками простейших масштабно-инвариантных иерархий являются следующие:

- а) ее члены распределены по уровням таким образом, что каждый член (за исключением тех, кто находится на самом нижнем уровне) имеет одинаковое количество N подчиненных;
- б) все подчиненные каждого уровня иерархии имеют одинаковых «вес» U, который равен весу непосредственного начальника, умноженному на коэффициент r<1. наиболее удобно рассматривать в качестве этого веса доход.

Первой попыткой рассмотреть данные социально-экономических процессов как самоподобную структуру является анализ Бенуа Мандельбротом цен на хлопок. Ученый исследовал статистику цен за период более ста лет и обнаружил, что колебания могут следовать скрытому математическому порядку во времени, который не описывается стандартными кривыми. Колебания цен в течение дня казались случайными, но Мандельброт смог выяснить тенденцию их изменения. Он проследил симметрию в длительных колебаниях цены и колебаниях кратковременных. Это открытие оказалось неожиданностью для экономистов. По сути, Бенуа Мандельброт применил для решения этой проблемы зачатки фрактального метода. В последствии для исследования временных рядов в рамках фрактального подхода Бенуа Мандельбротом был предложен R/S метод, базирующийся на исследованиях английского ученого Херста. Метод основан на анализе параметров размаха (наибольшим и наименьшем значением на изучаемом отрезке) и среднеквадратичного отклонения.

В российской практике впервые на фрактальную природу социальных данных ученые обратили внимание также в рамках изучения временных рядов. Так Давыдов А.А. [Фрактальный анализ социальных процессов] провел измерения фрактальной размерности множества социальных процессов: от социокультурных характеристик П.Сорокина («любовь», «индетерминизм», «идеализм», «реализм», «этика принципов») до данных мониторинга общественного мнения, проводимых ВЦИОМ за период 1994-2001 гг. (доверие Жириновскому, Зюганову, Явлинскому, градации «экономические реформы нужно прекратить», «все не так плохо и можно жить», «терпеть наше бедственное положение уже невозможно») и мониторингов общественного мнения жителей Западной Европы (доля очень счастливых, доля полностью удовлетворенных жизнью, доля респондентов, считающих, что следующий год будет хуже, чем предыдущий).

Сериков А.Е. [Фрактальный анализ временных рядов // Социология 4М. 2006. №22. С. 162-183], в свою очередь, проанализировал данные суммарного количества массовых политических событий и беспорядков (восстания, мятежы, бунты, войны, погромы и т.д.), собранные А.Л. Чижевским в 21 странах за период с марта 1923 г. по декабрь 1927 г. Было обнаружено, что распределения различий временного ряда на различных уровнях подобны друг другу (см. рис. 2).

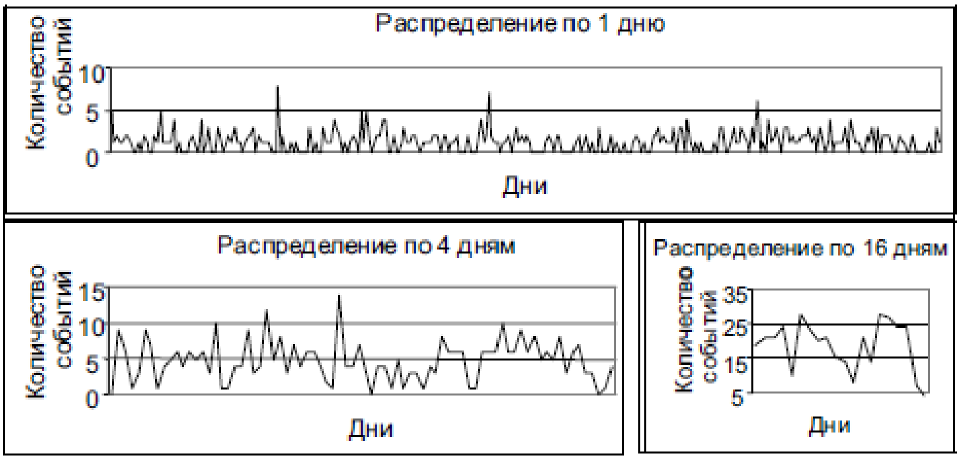


Рис. 2. Количество событий с участием масс в марте-декабре 1923 г.

Вычисление фрактальных размерностей данных временных рядов также подтвердило гипотезу о их фрактальной природе.

Наиболее широкое практическое применение фрактальный подход в анализе временных рядов получил в области анализа динамики курсов валют. Ярким преимуществом здесь является наличие критического значения фрактальной размерности кривой, там, где система теряет устойчивость и переходит в нестабильное состояние. Таким образом, можно определить временной отрезок с наиболее высокими рисками, даже предсказать финансовый кризис. Это можно заметить, если проанализировать динамику курса доллара в период кризиса 1998 года (см. рис. 3.) Непосредственно перед скачком курса хорошо заметен резкий подъем фрактальной размерности.

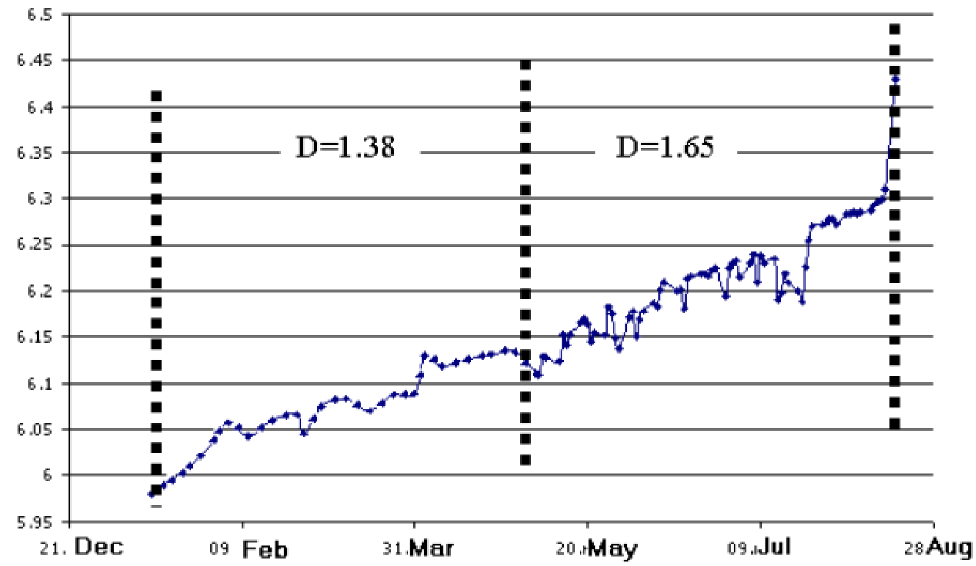


Рис. 3. Фрактальные размерности в динамике курса доллара к рублю в период финансового кризиса 1998 г.

То есть, оценку изменения фрактальной размерности можно использовать как некий индикатор приближающейся нестабильности и кризиса.

Однако возможности применения фрактальной теории для данных результатов социальных процессов не ограничивается лишь временными рядами. Как показывает практика, самоподобные структуры встречаются и в различных статических распределениях. Проиллюстрировать это можно примером распределения людей в городах и населенных пунктах, предложенным М.Шредером [Фракталы, хаос, степенные законы. С. 252]. Так, если некоторую обширную территорию, допустим, материк, разделить на две равные части, то доля населения на одной части составит p (причем, $p > 0.5$), а на другой – $1 - p$. Разумеется, размер избыточной доли p зависит от того, как была проведена линия, и в примере она проведена так, чтобы максимизировать величину p.

Если продолжить деление и разделить более густонаселенную часть еще пополам при сохранении условия максимизации численности в одной из частей, то окажется, что доли населения здесь составят p^2 и $p(1 - p)$. Если же разделить менее густонаселенную часть на две, то там доли населения составят, соответственно, $(1 - p)p$ и $(1 - p)^2$.

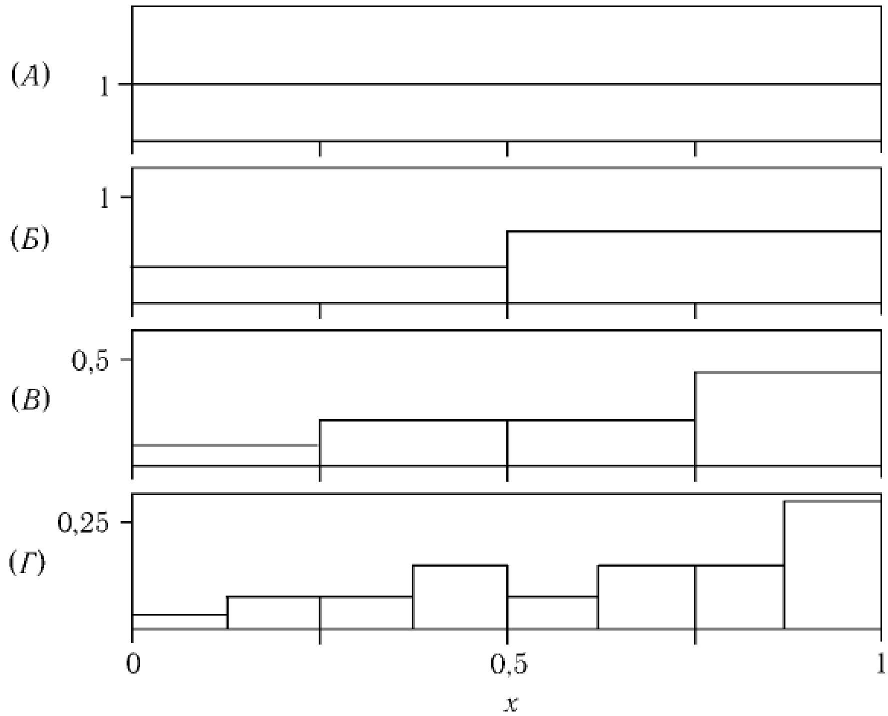


Рис. 4. Пример построения самоподобного распределения с помощью повторения операции деления пополам.

Итерация такого разделяющего процесса приводит к асимптотически самоафинному распределению (см. рис. 4). Мы начинаем с однородного распределения вероятностей в пределах единичного интервала (рис. 4А). После однократного деления интервала на две равные части находим две вероятности: $1 - p$ и p для двух половин единичного интервала. При $1 - p < p$ получается одноступенчатое распределение, представленное для $p = 2/3$ на рис. 4Б. Повторное деление приводит к четырем интервалам и распределению вероятностей, показанному на рис. 4В. Результат третьего деления – на рис. 4Г. В пределе, после бесконечного количества делений, получается самоафинное распределение: левая половина, растянутая с коэффициентом 2 в горизонтальном направлении и с коэффициентом *1(1 – p)* в вертикальном, воспроизводит распределение в целом.

При условии, что вероятность *p* = 0.7 характерна для населения земного шара, то после 18 делений всей суши останется два «отшельника», живущие на площади 576 км^2 = 24 км * 24 км в самом малонаселенном районе планеты, где-нибудь в Сибири, тогда как на территории такой же площади в современном мегаполисе будет находиться 8 млн. человек. Согласно такой простейшей модели, большинство людей (3,5 млрд. человек) живут в 60 000 населенных пунктов, численность жителей каждого из которых равна от 20 000 до 300 000 человек.

В настоящее время применение фрактального подхода распространено в анализе динамики рынка с применением в финансовой сфере, стратегическом планировании и управлении. За последнее несколько лет в российских средствах массовой информации появляются публикации, рассчитанные на широкого пользователя, в таких журналах, как например, «Менеджмент в России и за рубежом» [Воробьев А.Д. Использование фрактальной теории в стратегическом планировании и управлении // Использование фрактальной теории в стратегическом планировании и управлении. №1, 2006]. Различным аспектам фрактальной теории посвящены также разделы таких порталов, как портал Московского международного синергетического форума [http://www.synergetic.ru/fractal/], где можно найти статьи на такие темы, как «Познание как фрактальное блуждание в мире», «Человек Кликающий: фрактальные метаморфозы», «Самоорганизация фрактального способа освоения коммуникаций сложного мира и образование» [Тарасенко В.В.], «Нелинейная динамика российского рынка: фрактальный подход к устранению хаоса» [Пашутин С.Б.]. Фрактальную теорию также рассматривают также в рамках тетрасоциологии

[Математический проект «Фракталы социальной гармонии в «золотой тетрасоциологии»] (Тетрасоциология – четырехмерная постплюралистическая социология или макросоциологическая теория, рассматривающая общество и человека как органическое единство четырех сфер воспроизводства, стремящихся к гармонии. Тетрасоциология объединяет социальную статику - учение о ресурсах общества, динамику - учение о процессах воспроизводства, структуратику - учение о сферах как структурах общества и генетику - учение о стадиях социального развития. [Семашко Л.М. Социология для прагматиков. Санкт-Петербург, 1999]).

Таким образом, фрактальная теория в настоящее время находит свое отражение в двух областях: анализе временных рядов, либо на философском уровне (с точки зрения теории хаоса). Рассмотрение же статических распределений объектов в пространстве измеряемых признаков как фрактальных структур на эмпирическом уровне пока не распространено в российской и зарубежной практике, однако подобный анализ позволяет исследователю посмотреть на данные «под другим углом и увидеть скрытые закономерности, которые невозможно обнаружить при «стандартном» анализе.

1.2.2. Распределение данных в пространстве признаков

Принято считать, что распределение значений в пространстве признаков подчиняется закону о нормальном распределении, даже если распределение не нормальное - «Для любого унимодального симметричного распределения, даже если оно отличается от нормального, не менее 56% наблюдений будут попадать в промежуток ±1 стандартное отклонение от среднего арифметического значения, для ±3 стандартных отклонений внутри указанного интервала окажутся не менее 95% наблюдений» [Девятко, 1998]. Суть нормального распределения заключается в том, что наибольшее число значений признаков приближается к средней величине, а крайние значения встречаются довольно редко.

Опыт исследований показывает, что нормальное распределение встречается чаще всего в природе. В социологических и экономических данных в чистом виде оно бывает достаточно редко, чаще бывает логнормальное распределение, либо распределения с «тяжелыми» хвостами, «выбросами», пропущенными наблюдениями, являющимися «смесью» различных распределений [Давыдов, 1995, с. 131]. Однако даже при отклонении от нормального распределения его удобство и привлекательность в силу полноты теоретических исследований позволяет «использовать» его, в том числе, и в применении к социологическим данным.

Одними из видов распределения, которым, по мнению некоторых исследований, характеризуются экономические и социологические данные, является класс обобщенных гиперболических распределений. Обобщенное гиперболическое распределение – это непрерывное вероятностное распределение, определяемое как нормальная смесь дисперсии-среднего со смешивающей плотностью обобщённого обратного Гауссова распределения. Оно включает в себя распределение Стьюдента, распределение Лапласа, гиперболическое распределение и распределение variance-gamma. Обобщенное гиперболическое распределение позволяет учитывать ассиметричность данных и «тяжелые» хвосты. На гиперболическое распределение при анализе результатов социальных процессов нередко обращают внимание. Так, при разнесении социальных явлений на немасштабируемые и масштабируемые, первые подчиняются закону нормального распределения, тогда как для вторых характерно гиперболическое распределение [Талеб, 2009]. При этом под масштабируемостью здесь понимается способность некоторых результатов усиливаться, расти без эквивалентных затрат. Типичным примером такого явления может служить распространение компакт-дисков с записями пианиста. Эффект масштабирования проявляется в том, что пианист играет один раз, его записывают, а потом аудитория его слушателей может расширяться сколь угодно сильно без дополнительных усилий со стороны пианиста. Если пианист приобретает популярность и тираж дисков становится огромным, то мы сталкиваемся с эффектом, когда получаемый конечный результат оказывается совершенно не пропорционален исходным затратам труда музыканта. Разница в свойствах масштабируемых и немасштабируемых явлений детерминируется различием их природы. Так, если первые принадлежат миру природных феноменов, то вторые результат социальных процессов. Классическим примером немасштабируемого явления может служить вес людей, а масштабируемого – их доход.

Гиперболические распределения – ближайшие родственники фракталов. Очевидно, что если исключить из рассмотрения первые наиболее многочисленные классы, то оставшаяся часть будет представлять собой все то же ранговое распределение гиперболического вида, которое будет также хорошо описываться законом гиперболического типа. В этом смысле можно говорить о самоподобии этого распределения – усеченная часть распределения такая же как и все распределение вцелом. Естественно, что имеет смысл рассматривать такое распределение в контексте фрактального образования.

Возвращаясь к нормальному распределению, необходимо отметить, что в силу центральной предельной теоремы для формирования нормального распределения требуется сравнительно одинаковое влияние независимых случайных факторов. В качестве примера можно привести цитату из классического учебника по теории вероятностей. Условия центральной предельной теоремы: «по существу сводятся к требованию, чтобы влияние на сумму отдельных слагаемых было равномерно малым, т.е. чтобы в состав суммы не входили члены, явно преобладающие над совокупностью остальных по своему влиянию на рассеивание суммы. […] Неучтенные случайные факторы, создающие рассеяние, обычно характерны своей равномерностью и отсутствием среди них резко преобладающих» [Вентцель,

1964. С. 287]. Обычно же мы не наблюдаем такую ситуацию. Часто какие-то факторы влияют на распределение значений больше остальных, то есть, возникает доминирующий фактор. Здесь можно привести следующий пример: «На рынках валют существуют «самые главные» игроки – центральные банки. На рынках государственных облигаций крупнейшими и уникальными игроками являются государственные казначейства. Очевидно, что при таком распределении инвесторов нельзя исключать, что несколько наиболее крупных игроков могут действовать согласованно как в силу случайных причин - например, одинаковая оценка новой информации руководством нескольких крупнейших банков страны, так и в силу договоренности» [Буховец, Бирючинская, Кораблина].

Принимая существование доминирующего фактора, можно предположить, что его величина превосходит сумму всех остальных факторов. Такой фактор может действовать не постоянно, его величина и направление могут меняться во времени и пространстве, но в любой момент его воздействие значительно превосходит все остальные.

Рассмотрим некоторую величину, на которую оказывают влияние три фактора, например X, Y и Z. Тогда факт доминирования можно будет представить в следующей формулировке: среди факторов в любой момент найдется такой, что его величина превосходит сумму остальных, например – X -d-фактор, если $X>Y+Z$. Другими словами, неравенство треугольника не выполняется для величин факторов X, Y и Z. Заметим, что выполнение неравенства треугольника в этом случае означало бы, что среди указанных выше факторов доминирующего фактора нет, т.е. все факторы примерно одинаковы по своей величине и роли.

Роль неравенства треугольника в математических построениях общеизвестна, и нарушение этого неравенства приводит к отказу от изотропных метрических пространств и рассмотрению в качестве области изменения величины некоторой фрактальной структуры.

Говоря о гиперболических распределениях в социальных данных, нельзя не сказать о проблеме использования таких распределений при анализе. Проблема состоит в том, что эти распределения не имеют конечного первого центрального момента. Другими словами, при наличии такого распределения посчитать среднее по выборочным данным невозможно: оценка среднего значения (выборочное значение) не стабилизируется при увеличении объема выборки, растет вместе с объемом выборки. Такие случаи особенно настораживали, судя по описаниям специалистов в социальной психологии. Получается, что средние значения нельзя распространять на генеральную совокупность, и поэтому необходимы другие методы оценки.

Делая вывод из вышесказанного, можно заключить, что введение в рассмотрение фрактальных структур, во-первых, дает возможность более корректного и более основательного анализа структуры данных, так как мы отказываемся от априорного принятия нормального распределения и делаем попытку глубже понять саму суть распределения объектов в пространстве признаков.

Во-вторых, фрактальная структура допускает возможность существования доминирующего фактора, наличие которого не исключено в данных социальных процессов, но не предусматривается в силу условий центральной предельной теоремы.

И, в-третьих, рассмотрение данных как фрактальных образований позволяет избежать возможных неадекватных результатов, связанных с использованием классической вероятностной схемы, например, оценки средних значений.

С помощью фрактальных методов можно, как будет показано ниже, в конечном счете, описать любую естественную или искусственную «систему», состоящую из отдельных «элементов», самоподобно связанных между собой. Кроме того, приоритетными в системе должны являться не свойства элементов, а правила их соединения, т.е. структура.

1.3. Фрактальный подход при анализе и интерпретации данных типологического анализа

1.3.1. Типологический анализ и интерпретация его результатов

Типологический анализ – это метод, направленный на разбиение некоторой совокупности объектов на содержательно различающиеся группы. В основе современного типологического анализа лежит понятие о нечетких множествах, т.е. множествах, не имеющих четких границ, когда переход от принадлежности элементов множеству к непринадлежности их множеству происходит постепенно, не резко, т.е. элементы некоторой предметной области относятся к ней лишь с известной степенью принадлежности. Построение типологии на эмпирическом уровне проводится по обнаруженному и теоретически интерпретированному основанию [Абушенко В.Л. Энциклопедия социологии]. В этом его главное отличие от процесса классификации объектов, главной целью которой является установление определенной структуры порядка, установление взаимосвязей между объектами и определение места каждого элемента исследуемого множества в системе. То есть, типологизация – это, по сути, содержательно интерпретируемая классификация.

Процесс типологизации начинается с отбора признаков для классификации объектов и определения в этом наборе типообразующих признаков.

Обычно предполагается, что задание непосредственно измеряемого признака автоматически приведёт к разбиению совокупности на классы. Практически это эквивалентно утверждению о достаточности выбранного признака для построения классов. Такая ситуация может быть характерна для областей с хорошо развитой теорией. Однако чаще встречается ситуация, при которой выбранный признак является необходимым, но не достаточным, т.е. с помощью только одного признака не удаётся выделить достаточно однородные классы, хотя производить группировку по этому признаку можно. В этом случае выбранный для классификации признак дополняется множеством других признаков, т.е. осуществляется переход к многомерному основанию классификации. Практически имеет место попытка замены одного ненадежного признака совокупностью признаков, которая кажется более надежной.

Иногда в ходе построения классификации разделяют понятия индикатора и классификационного признака. Под индикатором в этом случае понимают те характеристики, которые непосредственно измеряются в ходе эксперимента. Как правило, это результаты операционализации теоретических понятий и представлений. Классификационными называются признаки, по которым непосредственно проводится классификация, т.е. те признаки, с которыми непосредственно работает алгоритм классификации. В ряде случаев таковыми являются сами исходные эмпирические индикаторы. Однако в рамках некоторых подходов классификационные признаки получают как интегральные показатели (индексы), например, средствами факторного анализа, т.е. рассматривают классификационные признаки как функции совокупности индикаторов.

Отбору признаков необходимо должен предшествовать содержательный анализ, проверка обоснованности, надежности используемых признаков. Этот этап является практически не формализуемым и полностью зависит от уровня теоретической оснащённости исследователя.

Общие требования, предъявляемые к исходным признакам, сводятся к тому, что они должны иметь одинаковый смысл для всех объектов, т.е. логическая структура описания и толкования (интерпретации) всех признаков должна быть одинаковой. Для отбора классификационных признаков обычно используются теоретические соображения, которые могут выступать в форме априорной классификации.

При этом следует иметь в виду, что признаки, несущественные для классификации, могут привести к искажению реальной структуры данных, выступать в роли «шума». Решение этой проблемы обычно производится посредством статистических процедур, в основе которых лежит оценка взаимосвязи признаков. Для этого могут использоваться такие, например, методы как корреляционный или факторный анализы. Применение этих процедур позволяет снизить количество рассматриваемых признаков. В первом случае за счёт отсева (отбраковки) дублирующих признаков, например, методом экстремальной группировки или корреляционных плеяд [Айвазян, Мхитарян, 1998, с. 567, с. 572], во втором случае используются методы построения новых показателей.

Отметим, что выбор признаков и метрики признакового пространства косвенно (неявно) отображает основание классификации и, в конечном счете, выбор самого алгоритма. Довольно часто основания выбора не указываются и уж подавно не аргументируются.

Переход от содержательной постановки задачи к ее математической формализации в большинстве случаев опирается на геометрическое представление данных о классифицируемых объектах. Суть этого подхода заключается в том, что каждый из $\boldsymbol{n}$ объектов заданной совокупности описывается одинаковым набором из $\boldsymbol{p}$ признаков. Значения этих признаков позволяют поставить в соответствие любому объекту точку $\boldsymbol{p}$-мерного пространства, каждая из координат которой соответствует определенному признаку. Пространство, в котором рассматриваемые объекты представлены точками, называется признаковым пространством.

В основу геометрического подхода положено предположение о том, что признаки и метрика выбраны таким образом, что близким в содержательном смысле объектам соответствуют близкие в смысле выбранной меры сходства точки признакового пространства. Понятие класса (кластера) как однородной совокупности точек признакового пространства отражает представление исследователя о тех свойствах, которыми должны обладать объекты, принадлежащие одному классу.

Математически формализация понятия однородности обычно заключается в задании некоторого критерия (функционала) качества разбиения. В этом случае задача кластерного анализа может быть сформулирована как задача отыскания оптимального с точки зрения выбранного критерия разбиения. Выбор функционала качества разбиения, - т.н. классификационного критерия, отражает представления исследователя о формировании совокупности эмпирических данных и связан с некоторыми гипотезами относительно характера неоднородности расположения точек в признаковом пространстве.

В качестве основных структурных гипотез рассмотрим следующие [Типология и классификация…, 1982]:

1. Гипотеза «компактности». Согласно этому предположению структура многомерных данных характеризуется наличием компактных кластеров, которые можно заключить в сферические или эллипсоидальные гиперповерхности. Компактность в данном случае понимается как близость объектов одного класса в признаковом пространстве к некоторому типичному представителю (центру, эталону, профилю), вокруг которого группируются все остальные объекты. Формирование такой структуры рассматривается как результат воздействия случайных факторов на реализацию эталонных объектов.

2. Гипотеза «связности» («непрерывности»). Ограничения, накладываемые на форму класса, кажутся обременительными и неоправданными в некоторых содержательных задачах. Более естественным в некоторых классификационных задачах выглядят требования связности объектов одного класса при одновременной изолированности различных классов. В этом случае наличие отчетливого разрыва в системе признаков, отделяющего кластеры друг от друга, является основой для формулировки классификационного критерия.

3. Гипотеза «унимодального распределения». Рассмотренные ранее структурные особенности многомерных данных в практических задачах часто бывают скрыты некоторым статистическим шумом, который в значительной степени ухудшает результаты кластерного анализа. В этом случае удобно ввести в рассмотрение функцию локальной плотности $f(x)$ и определить кластер как область признакового пространства, в которой эта функция унимодальна и имеет локальный максимум. Задача классификации в этом случае становится похожей на известную в статистике задачу разделения смесей, однако в отличие от последней в задаче классификации не требуется делать никаких предположений относительно функций распределения отдельных классов.

Интерпретация результатов применения метода анализа - это процесс приписывания содержательного смысла полученным в результате применения математического метода формальным символам, выражениям, отношениям и т.д. [Толстова, 1987]. В интерпретации необходимо соблюдение принципов:

Принцип согласования интерпретаций означает согласование интерпретаций результатов применения математического метода с интерпретацией данных. В первую очередь это учет в процессе интерпретации результатов применения математического метода той модели, которая заложена в используемом методе. Пример: если мы строим типологию изучаемых объектов с помощью метода классификации, направленного на поиск вытянутых сгущений (имеющих форму огурцов) в исходном признаковом пространстве, бессмысленно при интерпретации получившейся классификации отождествлять каждый класс с его центром тяжести (т.е. с объектом, координаты которого равны средним арифметическим значениям классификационных признаков). Такая интерпретация противоречит интерпретации исходных данных, в соответствии с которой искомые типы определяются закономерностями, обусловливающими вытянутый вид искомых классов, а отнюдь не средним уровнем рассматриваемых признаков. Центры тяжести разных классов могут быть близкими. Вычисление средних арифметических для одного класса может стать некорректным из-за большой растянутости отвечающего классу сгущения в признаковом пространстве и т.д.

- Принцип дополнения формализма состоит в том, чтобы в процессе интерпретации результатов применения математического метода обнаружились соображения, не отразившиеся в интерпретации исходных данных. Успех реализации этого принципа зависит от интуиции исследователя. Например, оценивая результаты классификации респондентов по бюджетам времени, необходимо учесть, что сходство объектов, как правило, определяется не только по их временным затратам, но и по способам времяпрепровождения..

1.3.2. Использование фрактального подхода при построении типологии

Возникающая проблема интерпретации классификационных разбиений решается чаще всего с помощью описания дескриптивных характеристик, а также построения моделей дисперсионного, факторного и регрессионного анализов (хотя обычно исследователи ограничиваются описанием средних значений показателей в каждом классе). В последнем случае выделенные классы (или – всю совокупность классов в целом) пытаются аппроксимировать некоторой функцией, чаще всего - линейной. Построенная регрессионная модель может быть использована как для анализа зависимостей одних признаков от других, так и для прогноза состояния системы. При этом мы предлагаем обратить особое внимание на то, что приближение объектов одного типа, а именно – совокупности изолированных точек признакового пространства, представляющих классифицируемые объекты, производится объектом другого типа – плоскостью. Мы полагаем, что было бы более естественно приближать одно множество точек другим множеством точек, особенно в том случае, когда свойства и характер поведения второго множества точек исследованы и представляются известными. В качестве аппроксимирующего множества мы предлагаем использовать фрактальные множества, которые получаются посредством выполнения итеративных процедур.

Приближение множества точек объектов признакового пространства сгенерированным с помощью процессов итерации множеством возможно при выявлении подобной структуры в самих изначальных данных.

Многими исследователями неоднократно отмечалось, что в практических задачах встречаются сравнительно немногие типы структурных неоднородностей. Так, в работе Е. Н. Князевой, С. П. Курдюмова (см. [Князева, 2001, с.10]) отмечается что, «немногие параметры порядка и немногие возможности, которые они имеют в принятии их индивидуальных состояний, отражают тот факт, что в сложных системах возможны только немногие определённые структуры которые, так сказать согласованы с поведением элементов. Иными словами, даже если некоторые конфигурации генерированы искусственно, извне, только некоторые из них действительно жизнеспособны….. Только определённые структуры из спектра потенциально возможных могут возникнуть, ибо они «разрешены» собственными свойствами системы, соответствуют им».

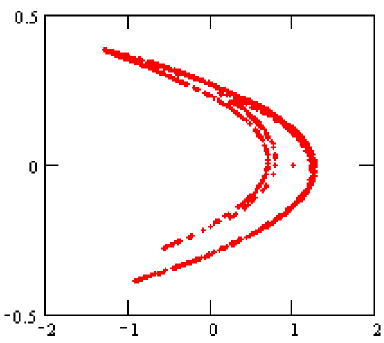
Поскольку распределение всех объектов в признаковом пространстве подчинено одному и тому же закону, то можно предположить, что классы, которые при этом обнаруживаются алгоритмами кластерного анализа, соответствуют областям устойчивости (т.е аттракторам системы итерированных функций) решения дифференциального уравнения. В качестве таковых в простейших случаях могут выступать устойчивый фокус, предельный цикл, странный аттрактор и др. Если полагать, как это традиционно делается в работах классификационной тематики, что классы соответствуют областям признакового пространства с высокой концентрацией объектов (точек), то в таком случае под классом следует понимать бассейн (область притяжения) особой точки устойчивого равновесия, т.е., другими словами, аттракторы (под аттрактором понимают множество, к которому стремится (притягивается) изображающая поведение динамической системы точка с течением времени [Ризниченко, 2003, с.29]) И, наоборот, области признакового пространства, не занятые объектами, будут представлять собой репеллеры, т.е. такие области, находясь в которых объект будет стремиться их покинуть в силу неустойчивости такого состояния.

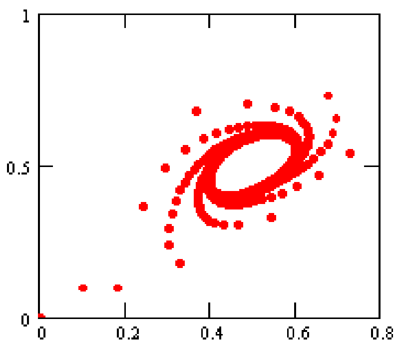
Действительно, сравнивая геометрические представления основных структур, выделяемых алгоритмами классификации, можно заметить, что гипотезе связности соответствуют классы, представленные в признаковом пространстве точками, расположение которых подобно аттрактору Хенона, изображённые на рис. 5А.

Гипотезе компактности будут соответствовать классы, образованные предельными циклами или аналогичные странному (стохастическому) аттрактору Лоренца. Примеры таких классов представлены на рис. 5Б.

Гипотеза унимодального множества хорошо представлена данными на рис. 5В, соответствующими бифуркации Хопфа. Тип аттрактора – предельная точка (устойчивый фокус), который в данном случае будет выступать в качестве моды класса.

Таким образом, структура многомерных данных, исследование которой является основной целью многомерной классификации, может моделироваться и хорошо воспроизводиться посредством представления разностных схем решений определённого класса дифференциальных уравнений или систем итерируемых функций.





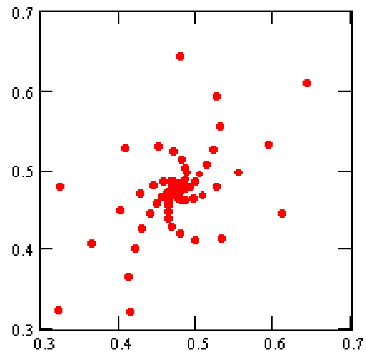


Рис. 5А. Аттрактор Хенона

Рис. 5Б. Предельный цикл

Рис. 5В. Фокус

Получаемое в результате применения иерархических алгоритмов классификационное дерево (дендрограмма), также может быть рассмотрено как некоторый мультифрактал (комплексный фрактал, который может детерминироваться не одним единственным алгоритмом построения, а несколькими последовательно сменяющими друг друга алгоритмами [Божокин, Паршин, 2001, с. 128]), в котором выделенные классы подобны всей совокупности.

1.3.3. Генерирование фрактальных структур системой рандомизированных итеративных функций

Развитие теории фракталов привлекло внимание к понятию структуры данных. Проблема исследования структурных особенностей многомерных данных возникала и раньше в задачах классификации или кластерного анализа. Однако отсутствие формальных критериев вызывало значительные трудности в сравнении различного рода структур.

В рамках фрактального подхода впервые была чётко сформулирована задача определения числовых характеристик структурных особенностей данных, делающих возможным, в частности, их сравнение и сопоставление. Кроме этого, в рамках теории фракталов был поставлен вопрос о генезисе (генерировании) самой структуры данных. Эти два аспекта реализовываются в рамках фрактального подхода.

В соответствие с основным принципом проверки статистических гипотез, можно считать, что нам что-то известно о структуре данных, если мы можем воспроизвести её. В задачах классификации обычно не рассматриваются механизмы формирования структуры данных. Одним из возможных подходов к описанию неоднородности многомерных данных будет подход, связанный с поиском и обнаружением фрактальных структур.

Предлагаемая в рамках данной работы процедура позволяет генерировать точки в признаковом пространстве таким образом, чтобы вновь полученные точки не нарушали бы фрактальной структуры уже имеющихся данных, т.е. удовлетворяли условию самоподобия. В качестве такой процедуры Буховцом А.Г. [Буховец, 2005] предлагается следующая схема действий:

Пусть в признаковом пространстве R^P определено некоторое множество точек $\mathbf{Z} = \{Z_i\}_{i=1}^{K}$, которое в дальнейшем будем называть протофракталом [Мандельброт, 2004].

1. В признаковом пространстве произвольным образом определяется точка $X_0 \in R^P$.
2. С помощью датчика случайных чисел определяется номер точки $Z_R \in \mathbf{Z}$ и вычисляются координаты новой точки $X_1 = (x_{11}, x_{12}, \ldots, x_{1P})$ по следующей формуле

$$x_{1,i} = \frac{x_{0,i} + (K-P) * \mu * z_{R,i}}{1 + (K-P) * \mu}, \quad (i = 1, \ldots, P) \qquad (1)$$

где K – число точек протофрактала;

P – размерность признакового пространства,

μ - некоторый масштабный коэффициент, характеризующий самоподобие структуры.

3. В дальнейшем точка X_1 принимается за исходную, а п. 2 повторяется столько раз, сколько точек необходимо получить. В результате применения такой процедуры будет получено множество точек $E_n = \{X_1, X_2, \ldots, X_n\}$, которое будем называть предфракталом.

Рассмотренная выше процедура может быть отнесена к классу рандомизированных итерированных функциональных систем, т.н. IFS (Iterated Function System [Кроновер, 2000, с.23]), в которых фракталы порождаются детерминированными правилами, выполняемыми случайным образом. В том случае, когда все преобразования точек являются аффинными сжатиями, соответствующее инвариантное множество называют самоафинным или L – системой (L – system) – системами Линдермайера. Полученное в результате выполнения этой процедуры множество называют аттрактором системы.

Рассмотрим основные свойства предложенной процедуры.

1. Неподвижные точки X^* этого преобразования можно определить, исходя из равенства:

$$X^* = \frac{X^* + kZ_R}{1+k},$$

где $k = (K - P)\mu$.

Очевидно, что в качестве неподвижной точки преобразования (при $\mu \neq 0$) может выступать любая точка протофрактала $\{Z_i\}_{i=1}^{n}$. Нетрудно видеть, что эти точки будут устойчивыми при выполнении условий: $K - P > 0, \mu > 0$.

2. Зависимость построенной структуры от значений параметра μ можно оценить, рассмотрев следующие соотношения:

$$\lim_{\mu \to 0} x_{L,i} = \lim_{\mu \to 0} \frac{x_{0,i} + (K - P) * \mu * z_{R,i}}{1 + (K - P) * \mu} = x_{0,i},$$

т.е. будет наблюдаться сходимость к одной точке.

В случае неограниченного увеличения значения параметра μ предельные точки будут совпадать с точками исходного протофрактала $\mathbf{Z} = \{Z_i\}_{i=1}^{K}$, т.к.

$$\lim_{\mu \to \infty} x_{L,i} = \lim_{\mu \to \infty} \frac{x_{0,i} + (K - P) * \mu * z_{R,i}}{1 + (K - P) * \mu} = z_{R,i}$$

При промежуточных значениях μ можно наблюдать, что каждая исходная точка становится основанием некоторой реплики, вокруг которой формируется группа точек, воспроизводящая структуру исходных данных. Причём, исходная точка, вокруг которой формируется новая структура, занимает в последней то же самое положение, какое она занимала в исходной. Другими словами, отчётливо наблюдается самоподобие генерируемой структуры. На рис. 6 представлена зависимость результатов выполнения процедуры от значений параметра *k*, а точнее, поскольку в задаче фиксированы значения *K* и *P*, то от значений параметра μ.

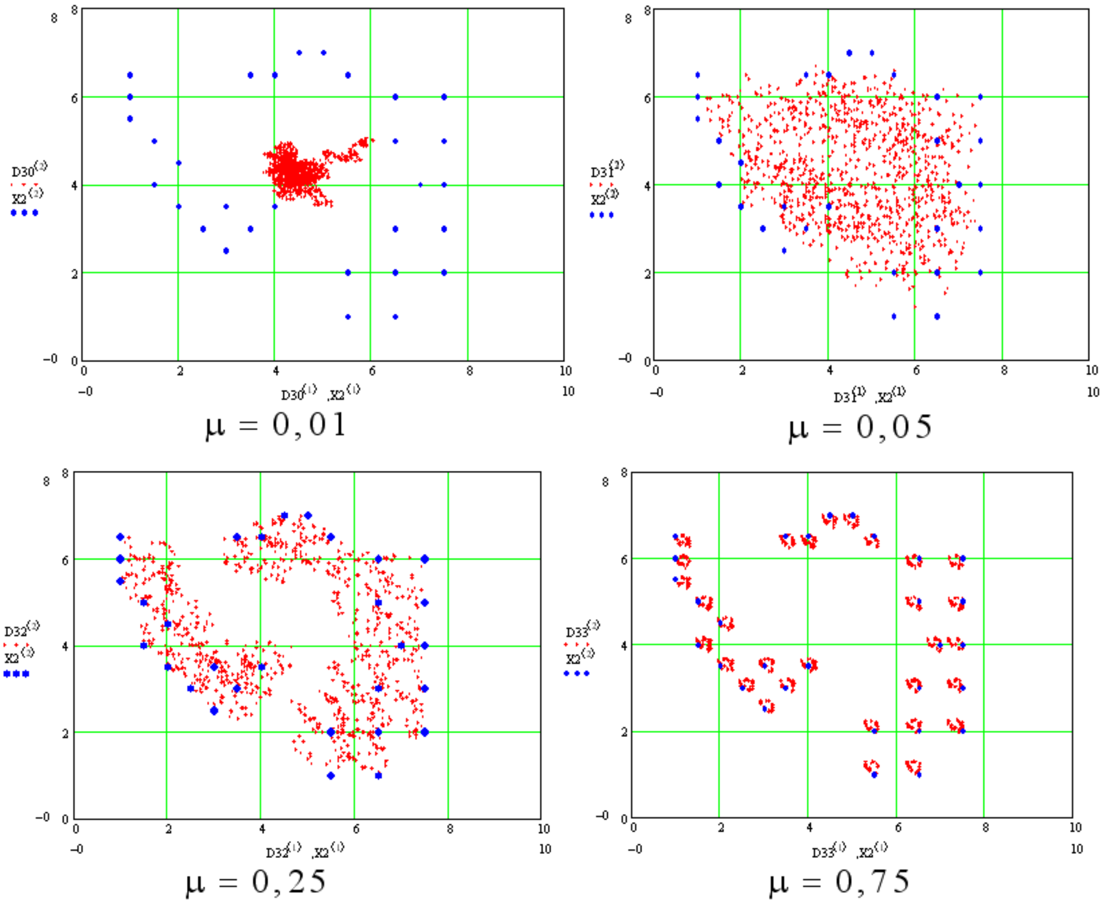


Рис. 6. Зависимость результатов выполнения процедуры генерирования фрактального множества от значения параметра μ.

Особенностью данного преобразования, как легко видеть из анализа (и это получило своё подтверждение в экспериментах), является то, что при все сгенерированные точки приближаются к одной неподвижной, а при все точки почти полностью совпадают с точками исходного массива. При промежуточных значениях можно наблюдать, что каждая исходная точка становится основанием некоторой реплики, вокруг которой формируется группа точек, воспроизводящая структуру исходных данных. Причём, исходная точка, вокруг которой формируется новая структура, занимает в последней то же самое положение, какое она занимала в исходной. Другими словами, отчётливо наблюдается самоподобие генерируемой структуры, что ещё раз подтверждает её фрактальность.

За счёт подбора параметра μ можно сделать так, чтобы описательные статистики, такие как среднее значение, минимальное и максимальное значения, выборочная дисперсия, матрица коэффициентов корреляции, исходной и генерируемой совокупностей не значительно отличались друг от друга. Как правило, именно при этом значении параметра особенно отчётливо выделяется структура многомерных данных.

Таким образом, в ходе реализации фрактального анализа для интерпретации результатов классификационных разбиений мы получаем точки протофрактала, сгенерированные по случайным точкам каждого класса. Результатом является матрица предфрактала, точки которой отражают структуру распределения данных.

Фрактальный анализ является средством математической интерпретации данных, в отличие от «традиционной» описательной интерпретации. В заключение сделаем еще одно замечание, касающееся понятия типичного объекта класса в классификационных задачах. Обычно, в качестве такого рассматривался объект, представленный соответствующей точкой, показатели которого были наиболее близкими к средним значениям выделенного класса. Такой подход может быть полностью оправдан в случае, когда кластер имеет распределение, близкое к нормальному. В случае выделения фрактальной структуры в качестве типичного объекта класса теперь будет выступать объект, занимающий в выделенном классе то место, какое он занимал бы в глобальной структуре генеральной совокупности. Так, например, если выделяются классы «бедных», «средних» и «богатых», то с точки зрения фрактального подхода типичным представителем класса «бедных» будет самый бедный, типичным «богатым» будет самый богатый, и только в среднем классе типичным будет оставаться представитель, характеристики которого наиболее близки к средним значениям выделенного класса.

Глава 2. Практическое применение фрактального подхода для интерпретации данных типологического анализа на примере типологии регионов России.

2.1. Обзор существующих работ по типологиям российских регионов и выбор оснований для классификации.

2.1.1. Обзор работ по типологиям российских регионов

Типологией регионов по уровню развития занимаются в России практически в течение всего XX века. Первые типологии посвящались исследованию экономического потенциала регионов России, однако они были лишены практического применения и имели чисто научное значение. В советское время была реализована так называемая районная «сетка» субъектов СССР: Центральный, Центрально-Черноземный, Северный и др. Ввиду того, что в Советском Союзе была распространена практика консервирования любых административных структур, эта районная сетка использовалась до распада СССР.

После распада СССР с расширением круга «потребителей» типологий усилился как теоретический, так и практический интерес к разработке методологии типологии регионов РФ.

Из теоретических типологий, не направленных на конкретное применение, можно выделить несколько основных. Одной из наиболее известных является типология регионов России по индексу развития человеческого потенциала (ИРЧП), построенная по методике ООН по принципу анализа ИРЧП в странах мира. Индекс развития человеческого потенциала был разработан в 1990 г. группой экономистов во главе с пакистанским экономистом Махбубом уль-Хаком (Mahbub ul-Haq) и представляет собой среднее арифметическое по трем индексам:

- индекс продолжительности жизни, или индекс долголетия $\frac{LE-25}{85-25}$

LE – ожидаемая продолжительность жизни. Упрощенно говоря, он обозначает среднее количество лет предстоящей жизни человека, достигшего данного возраста.

- индекс образования $\frac{2}{3} \times ALI + \frac{1}{3} \times GEI$

ALI – уровень грамотности взрослого населения, GEI – доля учащихся в возрасте 7-24 лет от общего количества трудоспособного населения в этом возрасте

- ВВП индекс $\frac{\ln(GDPpc) - \ln(100)}{\ln(40000) - \ln(100)}$

Расчет и применение индекса развития человеческого потенциала вызывает много критики, доля которой касается некорректности его применимости для классификации российских регионов. В частности, методика ПРООН рассчитывает уровень образования, исходя из двух параметров: уровень грамотности (с весом в 2\3) и доля учащихся трех ступеней образования в возрасте от 7 до 24 лет (с весом в 1\3). Однако для России такой способ не совсем адекватен, так как в условиях обязательного среднего образования уровень грамотности достигает почти 100% и примерно одинаков во всех регионах страны. То есть учет уровня грамотности с весом в 2\3 только сглаживает межрегиональную дифференциацию и не позволяет выявить корректную типологию субъектов. Недостатком данного подхода является игнорирование неполноты статистической информации: многие показатели могут быть получены лишь при всеобщей переписи населения, которая проводится раз в десять лет. Также некорректные данные по образованию обнаруживаются в Москве и

Санкт-Петербурге за счет того, что большая доля студентов, обучающихся в этих мегаполисах, проживают в близлежащих областных населенных пунктах.

Из общей критики индекса развития человеческого потенциала интересно отметить две вещи. Первая – это не совсем адекватные способы максимизации индекса. К примеру, чтобы максимизировать ИРЧП за счет образования, необходимо сделать всех трудоспособных жителей в возрасте 7-24 лет учениками и студентами, что нереально и, как минимум, неразумно. Второе, что делает индекс неадекватным, это то, что для страны бессмертных с бесконечным ВВП на душу населения ИРЧП будет равен 0,666, что ниже Таджикистана и Южной Африки.

Еще одна региональная классификация дана в проекте ТАСИС «Анализ развития регионов России (типология регионов, выводы и предложения)» [Экспертный институт, 1996]. В типологии использовано 11 основных показателей из 6 групп: общие – 2, демографические – 1, уровня жизни – 5, экономические – 1, финансовые – 1, структурные экономические – 1. Основной целью данной типологии был анализ адаптации регионов в условиях трансформации экономической системы, и поэтому, по мнению авторов, различия между регионами достаточно полно описываются показателями динамики промышленного производства и уровня доходов населения.

Типология регионов России по комплексу показателей здоровья населения и формирующих его факторов представлена в работе Мартынова А.С. и Виноградова В.Г. [Мартынов А.С., Виноградов В.Г. Медико-экологическая оценка условий жизни населения. Типология регионов России по комплексу показателей здоровья населения и формирующих его факторов., 1998] В этой работе медико-экологическая оценка условий жизни населения дана на основании типологии регионов по следующим 10 основным показателям из 5 групп: общие – 1, демографические – 3, социальные – 3, показатели уровня жизни – 1, характеризующие рынок труда

– 2. Среди показателей нет собственно экологических, но результатом работы является вывод об экологическом состоянии в регионах, сделанный не на основании эмпирических показателей и их интерпретаций, а на основании теоретических рассуждений, построенных как выводы из анализа эмпирических показателей. Индикаторами сложившейся медико-экологической ситуации выступают показатели здоровья и уровня жизни населения.

Еще одной работой в данной области, заслуживающей внимание, является работа «Типология российских регионов», выполненная в рамках проекта CEPRA (проект по сотрудничеству и техническому содействию, финансируемый Канадским правительственным агентством по международному развитию) в Институте Экономики Переходного Периода [Бутс Б., Дробышевский С., Кочеткова О., Мальгинов Г., Петров В., Федоров Г., Хехт А., Шеховцов А., Юдин А. Типология российских регионов. 2002 492 с.]. Типология получена на основе качественного анализа и совмещения результатов многомерной классификации регионов по показателям экономического потенциала, уровня жизни и инвестиционной активности. Инвестиционная активность представлена переменными:

- отношение инвестиций в основной капитал к валовому региональному продукту (ВРП), характеризующее абсолютный уровень инвестиций в регионе;

- относительные темпы роста инвестиций в основной капитал на региональном уровне по сравнению со среднероссийским уровнем, которые отражают межрегиональные различия в инвестиционной активности, а также межвременные предпочтения экономических агентов, действующих в регионе;

- отношение объема иностранных инвестиций к ВРП.

Экономический потенциал региона характеризуется следующим набором показателей:

- отношение темпов роста валового регионального продукта и ВВП России, характеризующее текущее экономическое положение в регионе относительно ситуации в экономике России в целом;

- уровень безработицы (отношение числа безработных к экономически активному населению), характеризующий как накопленное сокращение объема производства в регионе, так и процесс создания новых рабочих мест (производств) в регионе, а также ситуацию на рынке трудовых ресурсов (наличие достаточного числа свободных трудовых ресурсов для расширения экономической активности);

- доля топливной промышленности в объеме промышленного производства региона.

В свою очередь, уровень жизни в регионах характеризуется переменными:

- уровень абсолютной бедности населения региона, характеризуемой долей населения с доходами ниже регионального прожиточного минимума;

- уровень относительной бедности населения региона, характеризуемый отношением среднедушевого дохода его населения к региональному прожиточному минимуму;

- величина межрегионального перетока доходов, характеризуемую отношением среднедушевых расходов населения региона к региональному прожиточному минимуму.

Данная работа, является одной из наиболее подробных и полных, так как, во-первых, выбор основания для классификации в ней производится с учетом анализа достоинств и недостатков вышеописанных типологий, и, во-вторых, в ее рамках выполняется несколько классификаций на основе каждого исследуемого параметра (экономический потенциал, уровень жизни, инвестиционная активность).

Помимо теоретических типологий проводились также и прикладные, направленные на применение конкретных результатов на практике. Здесь можно выделить типологии, направленные на:

- оценку инвестиционного и предпринимательского климатов (на данный момент известно довольно много работ в этом направлении, где основанием для классификации служили от одного до нескольких десятков показателей; наиболее распространенный метод – ранжирование регионов по уровню их привлекательности [«Инвестиционный климат регионов России: опыт оценки и пути улучшения». М., ТПП РФ, «Альфа-Капитал», 1997; «Индексы инвестиционных рисков. Аналитический проект «Россия в третьем тысячелетии»». М., АО «Триада», 1994; Тихомирова И. «Инвестиционный климат в России: региональные риски». М.: Издатцентр, 1997; Акимов М. «Дорогая моя Русь (таблица инвестиционной привлекательности регионов России». – Профиль, 1997, №32; Котляр З. «Инвестиционная привлекательность регионов России». – Деловой мир, 15.09.1993; Nagaev S., Woergoetter A. «Regional Risk Rating in Russia». Vienna, Bank Austria, 1995; «Russian regions: Credit Suisse First Credit Ranking», 1998]). Пример одного из таких исследований – расчет инвестиционного рейтинга российских регионов по методике Марченко Г. и Мачульской О., публикуемый ежегодно в журнале «Эксперт». В основе методики лежат 9 основных групп показателей: общие (1 показатель), демографические (1), социальные (1), уровня жизни (1), характеризующие рынок труда (1), экономические (2), финансовые (5), транспорта и связи (3), инновационного потенциала (6). Итоговая типология составляется с помощью экспертного подхода. Данные исследования позволяют сделать некоторые выводы. Во-первых, оценка показателей инвестиционного потенциала и риска отражает высокую дифференциацию инвестиционных условий регионов. Во-вторых, вне конкуренции для инвесторов находятся Москва и Санкт-Петербург, обладающие максимальным потенциалом и минимальным риском. В-третьих, в первой десятки находятся почти все регионы – доноры федерального бюджета. В-четвертых, последние места по-прежнему занимают автономные округа и автономная область, слаборазвитые республики. В-пятых, значительно повышается рейтинг

- выявление динамики и специфики производства (здесь основанием для классификации служат структурные экономические и политические показатели). Примерами подобных исследований является типология динамики производства в регионах России, разработанная Институтом Экономики РАН [Маркова Н., Беденков А. «Социально-экономическое положение регионов России (обзор)». – Вопросы экономики, 1995, №3], где классификация была проведена с целью анализа факторов межрегиональной дифференциации по показателям динамики промышленного производства и рекомендаций механизмов государственной политики для разных групп регионов, и типология регионов России по показателям специализации хозяйства, осуществленная в рамках российско-канадского исследования проблем регионализма в России, представлена А. Галкиным и А. Казаковым [Galkin A., Kazakov A. «A typology of Russia's regions and the case study approach». Ch.2, 1998], где в качестве критериев экономической специализации регионов использовались как количественные показатели (структуры экономики регионов, объемов экспорта, его доли в суммарном объеме экспорта России и т.д.), так и некоторые качественные характеристики для описания социально-экономических явлений, характерных для регионов разных типов.

- формирование региональной и бюджетной политики (цель данных типологий – выявление кризисных территорий для оказания им экономической помощи в рамках региональной политики, следовательно, подход к отбору параметров для классификации имеет достаточно большое значение. Так, к примеру, в классификации кризисных территорий, подготовленной в Центре геополитических исследований института географии РАН [Бородулина Н. А. «О классификации кризисных территорий» (рабочие материалы). М., ИГ РАН, 1996] используется 48 показателей уровня жизни, здоровья населения, здравоохранения и состояния природной среды, образования и социальных условий образования).

Обзор данных исследований позволяет сделать несколько выводов, связанных с дальнейшей работой в направлении типологии российских регионов.

Во-первых, большинство исследований проводятся с той или иной степенью регулярности, что позволяет оценивать ситуацию в динамике. В свою очередь, из динамического анализа следует то, что состояние регионов России довольно нестабильно, и поэтому нельзя ограничиваться проведенными исследованиями, что делает вопрос о целесообразности проведения типологии российских регионов актуальным.

Во-вторых, как мы видим, задача типологического анализа регионов имеет практическое применение, даже если имеет место теоретическая типологизация (при разработке практических типологий, безусловно, изучаются теоретические). В этом случае, актуальным является усовершенствование методики типологизации.

И, в-третьих, отдельный интерес представляет сравнение оснований приведенных типологий. Мы видим, что некоторые показатели учитываются практически во всех типологизациях, даже достаточно узкоспециализированных. Эти выводы будут учтены при выборе основания для классификации регионов России, проведенной в рамках данной работы.

2.1.2. Выбор оснований для классификации

В рамках данной работы для иллюстрации возможностей использования фрактального анализа при интерпретации классификационных разбиений будет проведена классификация регионов России. Исходя из анализа основных теоретических работ, посвященных типологии российских регионов были выбраны следующие показатели:

- Уровень жизни в регионе);

- Инвестиционная активность;

- Экономический потенциал.

Уровень жизни является одним из наиболее сильных показателей дифференциации регионов России, так как после распада СССР и до сих пор наблюдается крайне высокий уровень межрегиональной дифференциации в уровне жизни населения [Типология российских регионов, с. 39]. В рамках проводимой автором классификации данный показатель включает в себя три индикатора:

- доля населения с доходами ниже прожиточного минимума (индикатор бедности);

- отношение среднедушевых доходов к прожиточному минимуму;

- отношение среднедушевых расходов к прожиточному минимуму.

Показатели среднедушевых доходов и расходов, безусловно, будут коррелировать между собой в большей или меньшей степени. Но, как это будет видно далее, их корреляция различается в кластерах, и поэтому включение в анализ показателя расходов является интересным с точки зрения изучения потребительского поведения, т.е. показатель дохода сам по себе – скорее экономический, а вместе с расходами получается изучение «движения денег» у людей с социологической точки зрения. В анализ не был взят показатель отношения среднедушевых доходов и расходов, так как это уже будет показателем финансового поведения населения (накопления), а это уже не показатель уровня жизни. Например, пенсионер получает пенсию 5000 руб., из них она на 2500 руб. тратит, а 2500 руб. откладывает. Получается, что ее сбережения – это половина ее дохода (высокий уровень накопления), но на самом деле и ее пенсия, и ее расходы – ниже прожиточного минимума, что является показателем низкого уровня жизни.

Инвестиционная активность – показатель не только текущей экономической активности региона, но и перспектив развития экономики как России в целом, так и отдельных ее областей. Однако построение единой инвестиционной функции для всей российской экономики чрезвычайно затруднено, или невозможно, не только из-за недостатков имеющихся данных, но и из-за различия типов инвестиционных процессов в различных субъектах РФ.

В рамках данной классификации инвестиционная активность измеряется следующими показателями:

- доля инвестиций в валовом региональном продукте;

- темпы роста инвестиций в 2009 г. по отношению к среднероссийскому уровню на соответствующий период 2008 г.

- отношение иностранных инвестиций к валовому региональному продукту.

Экономический потенциал показывает возможности развития региона с точки зрения внутренних «двигателей» развития (в отличие от инвестиционной активности как внешнего «двигателя») и включает в себя:

- уровень безработицы в регионе (доля безработных от экономически активного населения региона);

- темпы роста валового регионального продукта по отношению к ВВП.

Все показатели, кроме прожиточного минимума (2010 г.) и доли инвестиций в основной капитал (2007 г.), были рассчитаны за 2008 г. (ввиду того, что большинство из них имеют привязку к ВРП, а последнее значение ВРП есть только за 2008 г.).

2.2. Результаты классификации и их интерпретация.

2.2.1. Результаты классификации регионов

Классификация регионов проводилась с помощью кластерного анализа методом k-means в программном пакете SPSS. В анализ вошли 83 региона.

Первым шагом анализа является определение положения регионов в пространстве двух главных компонент (рассчитанных методом факторного анализа – см. приложение 1) – см. рис. 7.

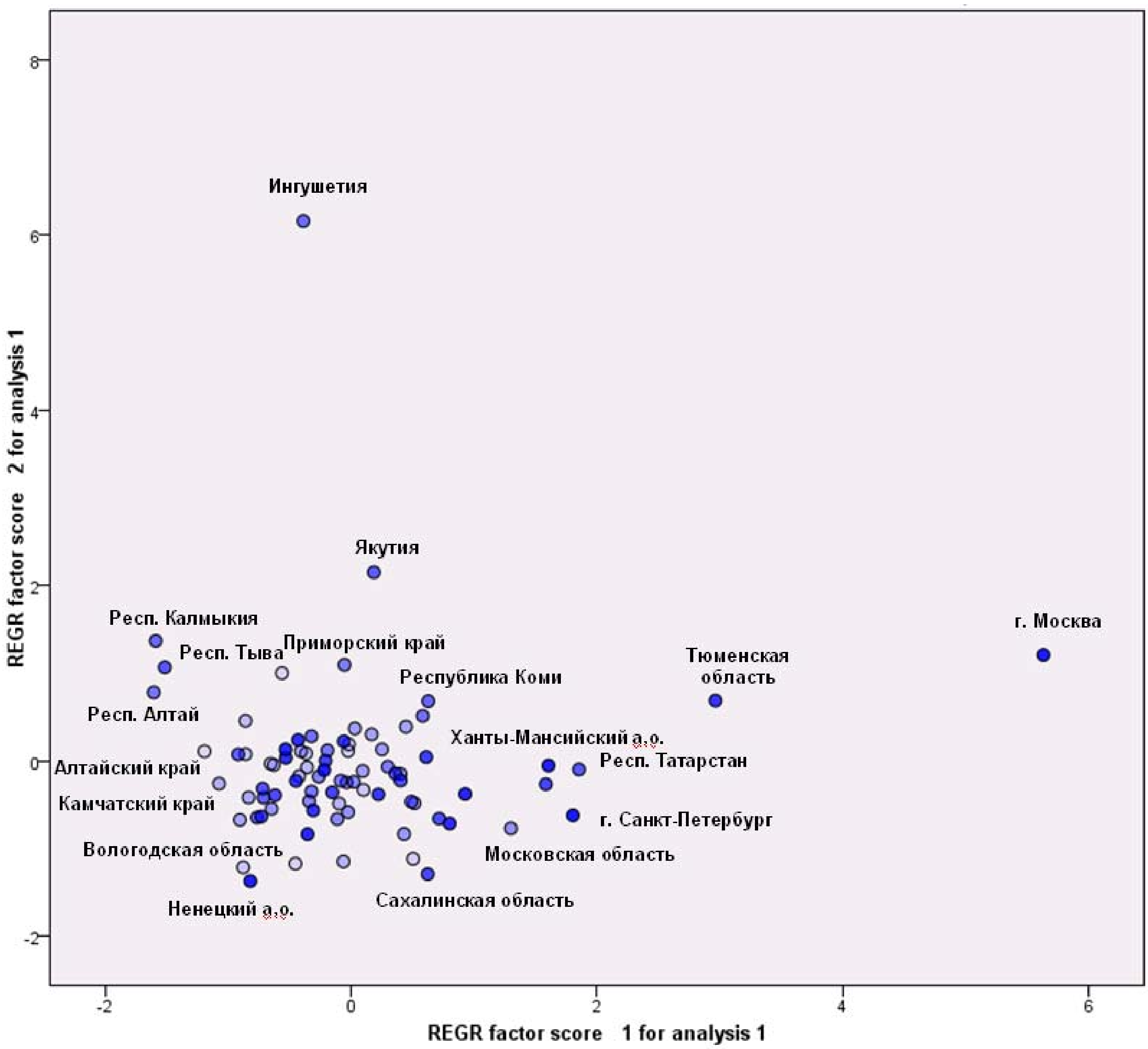


Рис. 7. Расположение регионов в пространстве двух главных компонент (без Чеченской республики).

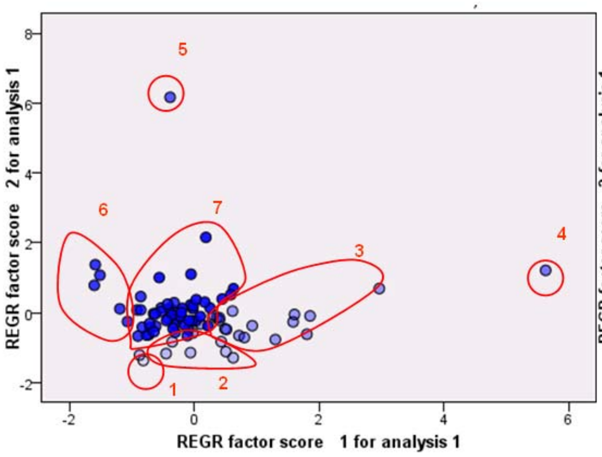
Далее было проанализировано различное число кластеров с точки зрения их содержательной наполненности. Так, при выделении 4-х и менее кластеров, получается один большой кластер, а в отдельные более мелкие выделяются явные «выбросы». Так, например, при выделении 4-х кластеров в состав 1-го входит 73 региона, в состав 2-го – 6 регионов, 3-го – 2 самых успешных (Москва и Тюменская область), 4-го – 2 самых депрессивных региона (Ингушетия и Чеченская республика).

При 5 кластерах получается разбиение на 2 крупных кластера (34 и 41 регион) и 3 мелких (в одном – 5 регионов, во втором – Ингушетия и Чеченская республика, в третьем – Москва).

При 6 кластерах получается уже 3 достаточно крупных кластера (31, 14 и 34 региона) и 3 – мелких (1-й – Москва, 2-й – Ненецкий автономный округ, 3-й – Ингушетия и Чеченская республика).

При 7-кластерном решении остаются 3 мелких кластера (1-й – Москва, 2-й – Ненецкий автономный округ, 3-й – Ингушетия и Чеченская республика), остальные же делятся на 4 кластера величиной в 7, 18, 34 и 20 регионов (см. рис. 8А)

8-кластерное решение похоже на 7-кластерное за одним лишь исключением, что в отдельный кластер выделяются республика Якутия, Приморский край и Амурская область (см. рис. 8Б).



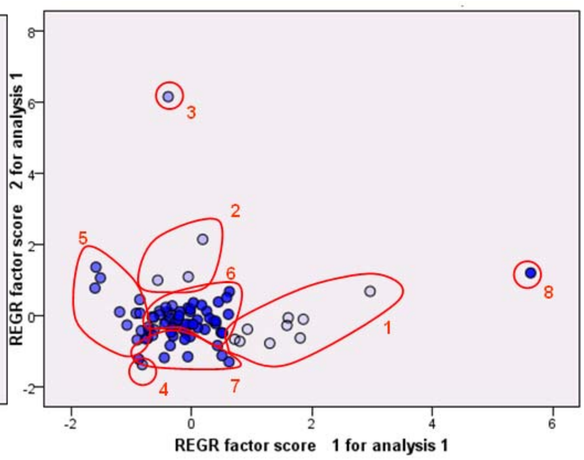


Рис. 8А. 7 кластеров в пространстве двух главных компонент.

Рис. 8Б. 8 кластеров в пространстве двух главных компонент.

При выделении большего числа кластеров происходит «отщепление» отдельных регионов (например, Сахалинской области, Чеченской республики и Ингушетии отдельно), поэтому их дальнейшее увеличение считается нецелесообразным.

При выборе между 7-кластерным и 8-кластерным решением предпочтение отдается 8-кластерному решению, так как структура распределения объектов в кластере, в состав которого входят республика Якутия, Приморский край и Амурская область (№2 на рис. 8Б), принципиально различается от кластера №6, а в 7-кластерном решении они рассматриваются вместе.

Итоговое решение содержит 8 кластеров (см. табл. 1).

Табл. 1. Средние значения по кластерам.

№ кластера	Кол-во регионов в кластере	Доля населения с доходами ниже прожиточного минимума	Отношение среднедушевого дохода к прожиточному мин.	Отношение среднедушевых расходов к прожиточному мин.	Темпы роста инвестиций в % к 2008 г. по сравнению со среднерос. уровнем, 2009	Доля инвестиций в основной капитал в ВРП_2007	Отношение иностранных инвестиций к ВРП	Отношение темпов роста ВРП и ВВП	Уровень безработицы
1	9	10,59	3,25	2,55	96,27	27,80	0,07	0,03	6,46
2	3	20,67	1,71	1,31	196,86	35,47	0,05	0,01	9,07
3	2	36,10	1,10	0,52	169,99	70,05	0,00	0,00	43,95
4	1	10,20	1,74	1,23	43,44	92,70	0,37	0,00	9,70
5	27	21,19	1,69	1,36	100,56	27,07	0,02	0,00	10,81
6	7	12,49	2,29	1,83	96,80	34,69	0,27	0,01	6,59
7	33	15,08	2,23	1,88	109,75	26,88	0,03	0,01	9,17
8	1	10,00	4,17	3,84	92,84	11,50	0,12	0,20	2,70

В состав 1-го кластера входят следующие регионы:

Московская область, г. Санкт-Петербург, республика Татарстан, Самарская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ, Ямало-Ненецкий автономный округ, Челябинская область.

Среднедушевые доходы и расходы в этих регионах практически самые высокие (выше только в Москве). Доля населения с доходами ниже прожиточного минимума минимальная (как в Москве и в Ненецком автономном округе). Т.е., уровень жизни определяется как высокий.

Темпы роста инвестиций чуть ниже, чем в целом по стране (на уровне 6 кластера). Доля инвестиций в основной капитал и отношение иностранных инвестиций в ВРП находятся на среднем уровне.

Экономический потенциал регионов 1-го кластера тоже высокий – находится на втором месте после Москвы.

Исходя из высокого уровня жизни и экономического потенциала данный кластер можно назвать **«самым благополучным».** Ближе всего к кластерному центру находится Свердловская область.

В состав 2-го кластера входят республика Якутия (Саха), Приморский край и Амурская область.

Этот кластер отличается очень самым высоким темпом роста инвестиций (197%). При этом, доля инвестиций в основной капитал в валовом региональном продукте ниже среднего (35,5%), иностранных инвестиций тоже немного (отношение иностранных инвестиций к ВРП равно 0,047, среднее же значение по России – 0,12). Уровень жизни невысокий, но и не самый низкий (доля населения с доходами ниже прожиточного минимума равна 20,7%; больше только у 3 и 5 кластера – «самых неблагополучных» и «неблагополучных» районов). Уровень безработицы здесь относительно низкий (9,07 против 13,68 по России). Темпов роста ВРП по отношению к ВВП не наблюдается (коэффициент равен 0,006).

Данные регионы можно определить как **«потенциально неблагополучные»** районы, которые быстрыми темпами приближаются к «неблагополучным». Самый яркий представитель – Амурская область.

В состав 3-го кластера входят республика Ингушетия и Чеченская республика.

Эти два региона являются **«самыми неблагополучными»**. Здесь самая высокая доля населения с доходами ниже прожиточного минимума – более трети населения (36,1%). Отношение среднедушевых доходов и расходов также самое низкое, причем отношение расходов к прожиточному минимуму меньше 1, т.е. люди не живут, а выживают. Темпы ростра инвестиций по сравнению со среднероссийским уровнем – 170%, т.е. чтобы жить, им требуется все больше и больше денег извне бюджета. Иностранные инвестиции в регионы отсутствуют, отношение ВРП и ВВП равно 0,001, уровень безработицы равен 43,95.

4-й кластер состоит из одного региона – **Ненецкого автономного округа**.

Он выделен в отдельный кластер из-за низких темпов роста инвестиций (по сравнению со среднероссийским – всего 43,44%) и высокой доли иностранных инвестиций (отношение иностранных инвестиций к ВРП самое высокое и равно 0,371).

В состав 5-го кластера входят следующие регионы:

Владимирская область, Воронежская область, Ивановская область, Костромская область, Рязанская область, республика Карелия, республика Калмыкия, Кабардино-Балкария, Карачаево-Черкессия, республика Марий Эл, республика Мордовия, Чувашская республика, Кировская область, Саратовская область, Ульяновская область, республика Алтай, республика Бурятия, республика Тыва, республика Хакасия, Алтайский край, Забайкальский край, Иркутская область, Томская область, Камчатский край, Хабаровский край, Магаданская область, Еврейская область.

Регионы 5-го кластера имеют довольно высокую долю населения с доходами ниже прожиточного минимума (21,2%), низкие среднедушевые доходы и расходы. Темпы роста инвестиций в основной бюджет по сравнению со среднероссийским уровнем невысокий – 100,56%. Уровень безработицы – 10,8. Если сравнивать положение регионов данного кластера с другим, можно заметить, что он практически идентичен со 2-м кластером («потенциально неблагополучные»), за тем лишь исключением, что во втором кластере наблюдается другая структура распределения данных (см. рис. 8А), и эта структура, видимо, обусловлена главным различием данных кластером – высокими темпами роста во втором кластере. Эти регионы определяются как **«неблагополучные»**. Самые яркие представители «неблагополучных» регионов – Забайкальский край, Чувашская республика и Ульяновская область.

В состав 6-го кластера входят регионы:

Чукотский автономный округ, Сахалинская область, Вологодская область, Архангельская область, Липецкая область, Калужская область, Белгородская область.

6-й кластер имеет относительно невысокую долю населения с доходами ниже прожиточного минимума (12,5%). Среднедушевые доходы и расходы выше, чем в среднем по России (отношение среднедушевого дохода к среднедушевых расходов к прожиточному минимуму равно 2,29 к 2 и 1,83 к 1,52 соответственно). Темпы роста инвестиций по сравнению со среднероссийским уровнем – 96,8%. Доля инвестиций в основной капитал ниже среднего – 34,69 (среднее значение – 44,95). 6-й кластер отличается от остальных самой высокой долей иностранных инвестиций (отношение иностранных инвестиций к ВРП равно 0,266 против среднего уровня в 0,12). Уровень безработицы в регионах 6-го кластера примерно равен уровню безработицы в 1-м кластере «самых благополучных» районов. Данные регионы можно определить как **«потенциально благополучные»**. Самый яркий пример «потенциально благополучных» регионов (находящийся ближе всего к кластерному центру) – Калужская область.

В состав 7-го – самого многочисленного кластера – входят регионы:

Брянская область, Курская область, Орловская область, Смоленская область, Тамбовская область, Тверская область, Тульская область, Ярославская область, республика Коми, Калининградская область, Ленинградская область, Мурманская область, Новгородская область, Псковская область, республика Адыгея, Краснодарский край, Астраханская область, Волгоградская область, Ростовская область, республика Дагестан, республика Северная Осетия-Алания, Ставропольский край, республика Башкортостан, Удмуртская республика, Пермский край, Нижегородская область, Оренбургская область, Пензенская область, Курганская область, Красноярский край, Кемеровская область, Новосибирская область, Омская область.

Регионы 7-го кластера занимают положение между «неблагополучными» и «потенциально благополучными» регионами, т.е. можно сказать, что это **«средние»** регионы без ярко выраженного потенциала перехода к неблагополучным или благополучным регионам. Среднедушевые доходы и расходы выше среднего уровня по России, но ниже, же у «благополучных» и «потенциально благополучных» регионов. Темпы роста инвестиций по сравнению со среднероссийским уровнем равны 109,75%. Доля инвестиций в основной капитал равна 26,88 (примерно как у «потенциально благополучных» и «благополучных», т.е. данные регионы – не реципиенты, они живут, в основном, за счет собственных средств). Доля иностранных инвестиций – как у «неблагополучных» регионов (0,029 против 0,025 соответственно). Уровень безработицы – как у «потенциально неблагополучных» и «неблагополучных» регионов. То есть, уровень жизни данных регионов чуть выше среднего, а экономический потенциал – чуть ниже среднего при низком уровне инвестиций. Самые яркие представители кластера (минимальное расстояние до кластерного центра) – Смоленская область и Пермский край.

В отдельный – 8-й – кластер выделен **г. Москва.** В Москве наблюдается самый высокий уровень жизни и экономического потенциала. Для сравнения: отношение среднедушевого дохода в Москве равно 4,17, а в «самых благополучных» регионах – 3,25. Отношение темпов роста ВРП и ВВП равно 0,204, а в «самых благополучных» регионах – 0,033. Уровень безработицы равен 2,7 против 6,46 у «самых благополучных» регионов. Что касается инвестиционной активности, в Москве наблюдается самая низкая доля инвестиций в основной капитал, однако количество иностранных инвестиций – не самое высокое (отношение иностранных инвестиций к величине валового регионального продукта равно 0,121, что ниже, чем у «потенциально благополучных» регионов, где оно равняется 0,371).

Таким образом, получившееся классификационное разбиение выявило:

1. Два больших кластера с низким потенциалом развития («неблагополучные» - 27 регионов – и «средние» - 33 региона).
2. Два кластера, регионы которых имеют положительный или отрицательный потенциал («потенциально неблагополучные» - 3 региона – и «потенциально благополучные» - 7 регионов).
3. Различие структуры расположения данных в кластерах. К примеру, «неблагополучные» и «средние» регионы образуют вполне однородные кластеры – рядом с кластерными центрами находится большая часть регионов (расстояние до кластерного центра 13 из 33 регионов «среднего» кластера и 11 из 27 регионов «неблагополучного» кластера меньше единицы). В то время как структура данных «потенциальных» кластеров не отличается подобной однородностью (расстояние до кластерного центра 3-х «потенциально благополучных» регионов равно 1,37, 1,48 и 1,73; у «потенциально благополучных» регионов Калужская область находится рядом с кластерным центром (расстояние равно 0,46),

Из выше сказанного можно сделать вывод о том, что изучение структуры данных при интерпретации данных классификационных разбиений должно играть главную роль, так как получаемые кластеры отличаются разной структурой, и эти различия в том числе объясняют содержательную наполненность кластеров. При стандартной интерпретации с помощью сравнения средних значений этот аспект не учитывается.

2.2.2. Реализация фрактального подхода в применении к классификации регионов России

По результатам классификации для выделенных 8-х классов была построена матрица А размеров *83x8* при значении μ = 17,8. Значение μ выбиралось экспертным образом (как уже было отмечено ранее, при $\mu \to 0$ все сгенерированные точки приближаются к одной неподвижной, а при $\mu \to \infty$ все точки полностью совпадают с точками исходного массива. При промежуточных значениях μ можно наблюдать, что каждая исходная точка становится центром, вокруг которого формируется группа точек, воспроизводящая структуру исходных данных. Причем, исходная точка, вокруг которой формируется новая структура, занимает в последней то же самое положение, какое она занимала в исходной. Другими словами, отчетливо наблюдается самоподобие генерируемой структуры, что еще раз подтверждает ее фрактальность [Буховец, 2006])

С помощью матрицы А были получены точки признакового пространства представляющие протофрактал Z. Результаты построения приводятся ниже.

$$
Z = \begin{pmatrix}
10.276 & 3.313 & 2.59 & 94.907 & 27.899 & 0.065 & 0.035 & 6.253 \\
20.638 & 1.688 & 1.289 & 200.858 & 34.995 & 0.046 & 0.006 & 8.455 \\
37.262 & 0.442 & 0.164 & 173.095 & 71.949 & -0.003 & -0.001 & 45.915 \\
9.955 & 1.709 & 1.191 & 39.852 & 96.355 & 0.389 & -0 & 9.738 \\
21.437 & 1.667 & 1.335 & 100.208 & 26.923 & 0.023 & 0.002 & 10.876 \\
12.33 & 2.296 & 1.827 & 96.794 & 35.277 & 0.279 & 0.007 & 6.367 \\
15.02 & 2.23 & 1.883 & 109.989 & 26.779 & 0.029 & 0.007 & 9.177 \\
9.695 & 4.28 & 3.953 & 91.895 & 10.636 & 0.125 & 0.211 & 2.321
\end{pmatrix}
$$

Каждая строка матрицы Z может быть соотнесена с ближайшим объектом классификации (см. приложение 1).

Точки протофрактала Z отражают структуру распределения объектов – регионов России – в пространстве выбранных для классификации признаков. Для сравнения их со средними значениями приводится таблица (см. табл. 2).

№ кластера	Доля населения с доходами ниже прожиточного минимума		Отношение среднедушевого дохода к прожиточному мин.		Отношение среднедушевых расходов к прожиточному мин.		Темпы роста инвестиций в % к 2008 г. по сравнению со среднерос. уровнем, 2009		Доля инвестиций в основной капитал в ВРП_2007		Отношение иностранных инвестиций к ВРП		Отношение темпов роста ВРП и ВВП		Уровень безработицы	
1	10,59	10,29	3,25	3,3	2,55	2,59	96,27	95,03	27,8	27,81	0,065	0,067	0,033	0,034	6,46	6,17
2	20,67	20,87	1,71	1,69	1,31	1,29	196,86	201,2	35,47	35,98	0,047	0,049	0,006	0,006	9,07	8,97
3	36,1	36,64	1,1	0,55	0,52	0,21	169,99	174,1	70,05	70,29	0	0,001	0,001	0,001	43,95	44,91
4	10,2	9,62	1,74	1,75	1,23	1,23	43,44	35,51	92,7	95,5	0,371	0,386	0,002	0,003	9,7	9,65
5	21,19	21,35	1,69	1,67	1,36	1,34	100,56	100,3	27,07	27,04	0,025	0,023	0,003	0,002	10,81	10,87
6	12,49	12,09	2,29	2,31	1,83	1,84	96,8	96,16	34,69	34,78	0,266	0,279	0,006	0,007	6,59	6,19
7	15,08	15,02	2,23	2,23	1,88	1,88	109,75	109,6	26,88	26,8	0,029	0,029	0,007	0,007	9,17	9,17
8	10	9,75	4,17	4,27	3,84	3,94	92,84	91,99	11,5	10,71	0,121	0,125	0,204	0,21	2,7	2,37

Табл. 2. Сравнение точек протофрактала со средними значениями по кластерам.

Исходя из табл. 1. можно сравнить средние значения показателей по каждому кластеру («обычная» интерпретация) с точками протофрактала Z (интерпретация с точки зрения фрактального подхода) и сделать выводы о том, насколько интерпретация средних значений не отражает структуру распределения исследуемых объектов в пространстве признаков.

Как правило, при интерпретации кластеров с помощью фрактального подхода получается ситуация, когда значения точек протофрактала, соответствующие признакам, в тех кластерах, где средние значения высокие, будет еще выше, а в тех – где они низкие – еще ниже. Так, к примеру, при разбиении населения по уровню дохода типичным представителем богатых людей будет самый богатый, бедных – самый бедный, и лишь среди людей со средним доходом самым ярким представителем будет человек со средним доходом [Буховец]. Однако, такая однозначная ситуация классификационного разбиения возникает не часто. Обычно кластеров бывает больше, и главной задачей аналитика при содержательной интерпретации является сравнение кластеров между собой, т.е. определение отношения кластеров друг к другу. В нашем случае использование фрактального подхода поможет оценить различия в структуре кластеров, обозначенные выше, а также обнаружить скрытые закономерности в общей структуре, и оценить, насколько «потенциальные» регионы отличаются от более однородных.

То есть, перед нами стоят две задачи:

1. Проанализировать положение «потенциально неблагополучных» и «потенциально благополучных» регионов ввиду обнаруженного различия их структуры.
2. Сравнить «благополучные» и «неблагополучные» регионы с целью выявления скрытых закономерностей в структуре данных.

Как мы видим из таблицы 2, точки протофрактала «потенциально неблагополучных» регионов различаются от средних значений не по всем показателям. Так, в показателе «уровень жизни» чуть выше доля населения с доходами ниже прожиточного минимума (20,87% – было 20,67%). При этом показатели отношения среднедушевого дохода и среднедушевых расходов к прожиточному минимуму остаются практически такими же (1,69 и 1,29, а было 1,71 и 1,31 соответственно). В инвестиционной активности различия обнаруживаются в показателе темпа роста инвестиций по сравнению со среднероссийским уровнем (среднее значение по кластеру 196,9%, 201,2% согласно протофракталу). Также немного больше оказывается доля инвестиций в основной капитал и доля иностранных инвестиций (35,98% и 0,049 – точки протофрактала и 35,47% и 0,047 – средние значения). Немного меньше становится уровень безработицы (8,97 – протофрактал и 9,07 – среднее значение). То есть, исходя из фрактального подхода, определение 2-го кластера как «потенциально неблагополучного» оказывается еще более оправдано, так как уровень жизни, исходя из структуры данных, оказывается ниже, чем если судить по среднему значению (регионы еще более «неблагоприятные»), а темпы роста инвестиций – роста помощи регионам, т.е. превращение их в реципиентов – еще больше.

В 5-м кластере – «потенциально благоприятных» регионов – доля населения с доходами ниже прожиточного минимума оказывается чуть меньше (12,09% против 12,49% среднего значения по кластеру). Из ярких различий можно отметить более низкий темп роста инвестиций по сравнению со среднероссийским уровнем (точка протофрактала – 96,16, среднее значение по кластеру – 96,8%), что означает склонность этих регионов к «донорскому» поведению. Также еще больше становится отношение иностранных инвестиций к ВРП (0,279 – точка протофрактала, 0,266 – среднее значение), хотя это в большей степени объясняется свойством фрактального подхода определять еще большие значения самым высоким показателям признаков (а отношение иностранных инвестиции в ВРП к 6-м кластере – самое высокое из всех кластеров).

Интересен еще тот момент, что показатели отношения среднедушевого дохода и среднедушевых расходов к прожиточному минимуму в точках протофрактала увеличивается, хоть и незначительно, в кластере «потенциально благополучных» регионов, в то время как в «среднем» кластере значения точек протофрактала не отличаются от средних значений ни по одному показателю, в том числе и по этим, что говорит о правильности классификационного разбиения с содержательной точки зрения (совпадение точек протофрактала со средними значениями по кластерам говорит, во-первых, о том, что объекты кластера действительно принимают «средние» значения, и, во-вторых, об однородности кластеров.

Второй задачей в применении фрактального подхода является сравнение «благополучных» и «неблагополучных» регионов с целью выявления скрытых закономерностей. «Неблагополучными» регионами в нашем случае являются объекты 2-го, 3-го и 5-го кластеров («потенциально неблагополучные», «самые неблагополучные» и «неблагополучные»). По анализу различий точек протофрактала и средних значений кластеров можно сделать следующие выводы:

1. Больше всего различий обнаруживается в кластере «самые неблагополучные» регионы. Такая ситуация возникает из-за того, что два объекта, входящие в состав кластера (Ингушетия и Чеченская республика) находятся достаточно далеко от остальных регионов России (см. рис. 8Б), т.е. подобные различия помимо характеристики самого кластера дают нам сведения о структуре всей совокупности данных.
2. Меньше всего различий обнаруживается в самом однородном кластере «неблагополучных» регионов, что говорит возможностях фрактального подхода «видеть» также и структуру расположения данных в многомерном пространстве признаков в рамках каждого кластера.

Из анализа «благополучных» регионов России можно следует, что «самые благополучные» регионы и Москва по уровню жизни и экономическому потенциалу еще больше вырываются вперед. Отношение иностранных инвестиций остается на прежнем уровне. Доля инвестиций в основной капитал оказывается чуть меньше в Москве (точка протофрактала – 10,71, измеренное значение – 11,5). Темпы роста инвестиций оказываются чуть ниже (точки протофрактала: 91,99 в Москве и 95,03 в «самых благополучных» регионах; средние значения: 92,84 в Москве и 96,27 в «самых благополучных» регионах). Подобные идентичные различия говорят о подобной структуре этих объектов в многомерном пространстве. То есть, можно предположить, что структура данных двух кластеров (Москвы и «самых благополучных» регионов) является увеличенным самоподобным множеством каждого из этих двух кластеров (безусловно, здесь речь может идти только о приближенном самоподобии).

В рамках фрактального подхода определенный интерес представляет рассмотрение 4-го кластера, в состав которого входит один объект – Ненецкий автономный округ. Его выделение в отдельный кластер неоднозначно (в отличие от Москвы, где все показатели кардинально отличаются от всей России). Если смотреть его расположение на двумерном пространстве, образованном двумя главными компонентами (см. рис. 8Б), то он «лежит» совсем рядом со вторым кластером («потенциально неблагополучные» регионы). Отличается от других кластеров он большим количеством иностранных инвестиций и низкими темпами роста инвестиций по сравнению со всей Россией. Подробное изучение матрицы точек протофрактала позволило получить некоторое возможное объяснение. Оказывается, что различия точек протофрактала от реальных значений параметров в Ненецком автономном округе очень похоже на различия, наблюдаемые в Москве – втором кластере, образованном из единственного региона. Так, доля населения с доходами ниже прожиточного минимума чуть ниже, чем есть на самом деле (точки протофрактала: 9,62 в Ненецком автономном округе и 9,75 в Москве; значения: 10,2 в Ненецком автономном округе и 10 в Москве). Отношение среднедушевого дохода к прожиточному мнению увеличено незначительно (точки протофрактала: 1,75 в Ненецком автономном округе и 4,27 в Москве; значения: 1,74 в Ненецком автономном округе и 4,17 в Москве). Темпы роста инвестиций уменьшаются (в Ненецком автономном округе различия в данном случае гораздо более значительны, чем в Москве – точки протофрактала: 35,51в Ненецком автономном округе и 91,99 в Москве; реальные значения: 43,44% в Ненецком автономном округе и 92,84% в Москве). Уровень безработицы и там, и там оказывается ниже (точки протофрактала: 8,97 в Ненецком автономном округе и 2,37 в Москве; значения: 9,07 в Ненецком автономном округе и 2,7 в Москве).

Ближайшие к точкам протофрактала объекты кластеров в некоторых случаях отличаются от объектов, ближайших к кластерным центрам. Так, самым ярким представителем «самых благополучных» регионов с точки зрения фрактального подхода является г. Санкт-Петербург (была Свердловская область), «потенциально благополучных» - Приморский край (была Амурская область), «неблагополучных» - республика Хакасия и Чувашская республика (был Забайкальский край и Чувашская республика). В свою очередь, в кластерах «потенциально благополучных» и «средних» регионов самые яркие представители не изменились (Калужская и Смоленская области соответственно).

2.2.3. Выводы и направления дальнейшей работы

В ходе практического применения фрактального подхода к анализу результатов типологического анализа было проиллюстрировано, каким образом с помощью генерирования фрактального множества можно «увидеть» структуру данных. Так, во-первых, мы видим различие в структуре данных в многомерном пространстве, по которым можем делать выводы о тех или иных особенностях отдельных кластеров. И, во-вторых, мы можем оценить общую структуру данных, выявить и проанализировать выбросы.

Такой способ обнаружения структуры данных очень ценен для социолога-исследователя, так как позволяет ему избежать необходимости отображения кластеров в пространстве двух главных компонент с потерей большей части информации только для того, чтобы оценить форму и структуру кластера.

Автором видятся очень широкие горизонты для дальнейшей деятельности в данном направлении. Так, одним из направлений является разработка методологии обнаружения характера расположения объектов в кластерах по точкам протофрактала (однородные / неоднородные, вытянутые и т.п.).

В целом, применение фрактальной теории в анализе данных дает очень большие возможности для изучения структуры данных и поиска закономерностей, которые невозможно выявить существующими на сегодняшний день методами, что делает ценным и актуальным продолжение работы в данной области.

Заключение

В данной работе основное внимание было сосредоточено на генерировании фрактальной структуры на основе реального расположения объектов в многомерном пространстве признаков.

В то же время использование фрактальной парадигмы открывает большие возможности при моделировании социальных процессов в случае т.н. качественных исследований, как на уровне построении теорий среднего уровня, так и при моделировании непосредственно самих процессов, в которых проявление самоподобия (например, при построении вертикальной структуры власти) заметны уже давно.

В любом случае, было бы методологически неоправданно не замечать появления новых парадигм в современных методах математического моделирования. Возможно, что те трудности, с которыми сталкивается современная методологическая (сторона) исследований как раз и связана с тем, что аппарат евклидовой геометрии (и статистики, как это уже было отмечено, например Хайтуном) не в полной мере соответствует именно природе социологических данных. С другой стороны, с помощью современного методного арсенала социолога не всегда возможно решить поставленные задачи (даже если не затрагивать тему евклидовой геометрии и теории вероятности). Как было показано в рамках практической части данной работы, использование фрактального подхода позволяет «увидеть» структуру данных на разных уровнях – как на уровне отдельных кластеров, так и на уровне всего массива данных.

Теоретическое значение фрактального анализа можно связать с решением одного из важнейших вопросов в теоретической социологии – вопроса о соотнесении социальных процессов на микроуровне и макроуровне. Возможны два радикальных ответа: первый заключается в том, что микроуровень и макроуровень социальной реальности никак между собой не связаны и подчиняются разным закономерностям; второй – в том, что социальные процессы едины на всех уровнях. Конечно, существует множество промежуточных вариантов.

В заключение данного раздела отметим следующее. Основатель социологии О.Конт [39,с.8] считал, что «Наша подлинная задача состоит в том, чтобы тщательно анализировать условия, в которых происходят явления, и связать их друг с другом естественными отношениями последовательности и подобия».

Список литературы

1. Anderberg M.R. Cluster Analysis for Applications. – N.Y.: Academic Press, 1973.
2. Hartigan. J. Clustering Algorithms. – N.Y.: John Wiley, 1975.
3. Lorr M. Cluster Analysis for Social Sciences. – San Francisco, Jossey-Bass, 1983.
4. Божокин С. В., Паршин Д. А. Фракталы и мультифракталы. — Ижевск: «РХД», 2001.
5. Буховец А.Г. Математическое моделирование структур многомерных данных в классификационных задачах, 2006
6. Буховец А.Г., Бирючинская Т.Я., Буховец Е.А. Использование фрактальных моделей в задачах классификации. - Системы управления и информационные технологии. 2009, 3.1(37), С. 117-121.
7. Буховец А.Г. Бирючинская Т.Я. Кораблина Н.А. Модели, учитывающие влияние доминирующего фактора. Экономическое прогнозирование: модели и методы. Материалы VI Международной научно – практиче-ской конференции 6 апреля 2010 г. – Воронеж: ВГУ, 2010. – ч.1, с. 61 – 66.
8. Буховец А.Г. Об одном подходе к задаче классификации // Социология: методология, методы, математические модели. 2004. № 18. С.82-105.
9. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1964. – 576 с.
10. Волгин Л. И. Гиперболическое распределение и его практическое применение.
11. Волков Ю.Г., Мостовая И.В. Социология: Учебник для вузов / под ред. проф. В.И. Добренькова. – М.: Гардарика, 1998. – 244 с.

12. Давыдов А.А. Анализ одномерных частотных распределений в социологии: эволюция подходов. // Социологические исследования. 1995. № 5. С. 131-137.
13. Девятко И.Ф. Методы социологического исследования. - Екатеринбург: Изд-во Урал, ун-та, 1998.- 208 с.
14. Дюран Б., Оделл П. Кластерный анализ / Пер с англ.- М.:Статистика, 1977. – 128 с.
15. Интерпретация и анализ данных в социологических исследованиях. – М.: Наука, 1987. – 255 с.
16. Истигечева Е.В. Оценивание параметров гиперболического и обратного гауссовского распределений. // Известия Томского политехнического университета. 2006. Том 309, вып. 6. с. 11-13.
17. Крыштановский А.О. Анализ социологических данных. – М.: Изд. Дом ГУ-ВШЭ, 2006. 281 с.
18. Кроновер Р.М. Фракталы и хаос в динамических системах. – М.: ТЕХНОСФЕРА, 2006. – 488 с.
19. Мандельброт Б. Фракталы, случай и финансы.– М.–Ижевск, 2004, 256 с.
20. Математические методы анализа и интерпретация социологических данных. – М.: Наука, 1989. – 173 с.
21. Талеб Н. Черный лебедь. – М.: Колибри, 2009. 528 с.
22. Татарова Г.Г. Основы типологического анализа в социологических исследованиях. – М.: Издательский Дом «Высшее образование и наука», 2007. – 236 с.
23. Татарова Г.Г. Типологический анализ в социологии – М.: Наука, 1993. – 103 с.
24. Типология и классификация в социологических исследованиях. М.: Наука, 1982. – 296с
25. Толстова Ю.Н. Измерение в социологии: Курс лекций. - М.: ИНФРА-М, 1998 - 224 С.

26. Факторный, дискриминантный и кластерный анализ: Пер. с англ./ Под ред. И. С. Енюкова. – М.: Финансы и статистика. 1989. – 215с.: ил.
27. Шредер М. Фракталы, хаос, степенные законы. – Ижевск, Удмурдский унивеситет, 2000.
28. Хайтун С.Д. Наукометрия. Состояние и перспективы. – М.: Наука, 1983, - 344 с.
29. Хайтун С.Д. Проблемы количественного анализа науки. М., 1989.
30. Чупахин И.Я., Бродский И.Н. Формальная логика // Издательство ЛГУ, 1977.
31. Шрейдер Ю.А., Шаров А.А. Системы и модели. М.: Радио и связь, 1982. – 152 с., ил.

Приложение 1. Предфрактал А.

A =

	1	2	3	4	5	6	7	8
1	0	0	$1.21\cdot10^{-8}$	0	$2.221\cdot10^{-3}$	0.952	0.046	0
2	$2.504\cdot10^{-7}$	0	$3.183\cdot10^{-15}$	0	0.048	$5.844\cdot10^{-10}$	0.952	0
3	0	0.046	$2.221\cdot10^{-3}$	0	0.952	0	$5.446\cdot10^{-6}$	0
4	0	0	0	0	0.954	0	0.046	0
5	0	0	0	0	0.998	$2.221\cdot10^{-3}$	$2.625\cdot10^{-7}$	0
6	0	$1.364\cdot10^{-12}$	0	0	0.046	0.952	$2.51\cdot10^{-7}$	$2.221\cdot10^{-3}$
7	$1.364\cdot10^{-12}$	$5.183\cdot10^{-6}$	0	0	0.998	0	$2.221\cdot10^{-3}$	0
8	$2.221\cdot10^{-3}$	0	0	0	$5.195\cdot10^{-6}$	$2.504\cdot10^{-7}$	0.998	0
9	$1.364\cdot10^{-12}$	$1.21\cdot10^{-8}$	0	0	$2.504\cdot10^{-7}$	0.952	0.048	0
10	0.998	0	0	0	$1.073\cdot10^{-4}$	$1.21\cdot10^{-8}$	$2.226\cdot10^{-3}$	0
11	$1.268\cdot10^{-8}$	0	0	$3.183\cdot10^{-15}$	$1.125\cdot10^{-4}$	0	1	0
12	$1.125\cdot10^{-4}$	0	0	0	0.998	$2.221\cdot10^{-3}$	$1.21\cdot10^{-8}$	0
13	$2.504\cdot10^{-7}$	$5.183\cdot10^{-6}$	0	0	$1.073\cdot10^{-4}$	$5.844\cdot10^{-10}$	1	0
14	0	0	0	0	0.048	$2.823\cdot10^{-11}$	0.952	0
15	$5.183\cdot10^{-6}$	0	0	0	0.046	$2.823\cdot10^{-11}$	0.954	0
16	$1.073\cdot10^{-4}$	$1.21\cdot10^{-8}$	$3.183\cdot10^{-15}$	0	0.046	0	0.954	0
17	$1.21\cdot10^{-8}$	0	$1.364\cdot10^{-12}$	0	0.046	$2.221\cdot10^{-3}$	0.952	0
18	$1.21\cdot10^{-8}$	0	0	0	$5.434\cdot10^{-6}$	$1.073\cdot10^{-4}$	0.048	0.952
19	$2.51\cdot10^{-7}$	0.046	0	0	0.952	0	$2.328\cdot10^{-3}$	0
20	$1.21\cdot10^{-8}$	0	0	0	0.048	0	0.952	0
21	$5.183\cdot10^{-6}$	0	0	0	0.046	0.952	$2.221\cdot10^{-3}$	0
22	0	0.046	$2.221\cdot10^{-3}$	0.952	$5.183\cdot10^{-6}$	$1.073\cdot10^{-4}$	$6.141\cdot10^{-10}$	0
23	$6.589\cdot10^{-14}$	0	$1.21\cdot10^{-8}$	0	0.048	0.952	$2.504\cdot10^{-7}$	0
24	0.046	$5.183\cdot10^{-6}$	0	0	$2.51\cdot10^{-7}$	0	0.954	0
25	$1.073\cdot10^{-4}$	0.046	0	0	$1.271\cdot10^{-8}$	$5.434\cdot10^{-6}$	0.954	0
26	$2.221\cdot10^{-3}$	0	0	0	$1.125\cdot10^{-4}$	$2.504\cdot10^{-7}$	0.998	0
27	0	0	0	0	0.046	$1.21\cdot10^{-8}$	0.954	0
28	0.046	0	$2.221\cdot10^{-3}$	0	$6.14\cdot10^{-10}$	0	0.952	0
29	0.952	0	$5.183\cdot10^{-6}$	0	0.046	$1.21\cdot10^{-8}$	$2.329\cdot10^{-3}$	0
30	0.046	0	0	0	$2.221\cdot10^{-3}$	$2.823\cdot10^{-11}$	0.952	0
31	$2.334\cdot10^{-3}$	$3.183\cdot10^{-15}$	0	0	0.952	0	0.046	0
32	$1.073\cdot10^{-4}$	0	0	0	0.046	0	0.954	0
33	0.046	0	0	0	$3.191\cdot10^{-15}$	$2.823\cdot10^{-11}$	0.954	0

A =

	1	2	3	4	5	6	7	8
1	0	0	$1.21\cdot10^{-8}$	0	$2.221\cdot10^{-3}$	0.952	0.046	0
2	$2.504\cdot10^{-7}$	0	$3.183\cdot10^{-15}$	0	0.048	$5.844\cdot10^{-10}$	0.952	0
3	0	0.046	$2.221\cdot10^{-3}$	0	0.952	0	$5.446\cdot10^{-6}$	0
4	0	0	0	0	0.954	0	0.046	0
5	0	0	0	0	0.998	$2.221\cdot10^{-3}$	$2.625\cdot10^{-7}$	0
6	0	$1.364\cdot10^{-12}$	0	0	0.046	0.952	$2.51\cdot10^{-7}$	$2.221\cdot10^{-3}$
7	$1.364\cdot10^{-12}$	$5.183\cdot10^{-6}$	0	0	0.998	0	$2.221\cdot10^{-3}$	0
8	$2.221\cdot10^{-3}$	0	0	0	$5.195\cdot10^{-6}$	$2.504\cdot10^{-7}$	0.998	0
9	$1.364\cdot10^{-12}$	$1.21\cdot10^{-8}$	0	0	$2.504\cdot10^{-7}$	0.952	0.048	0
10	0.998	0	0	0	$1.073\cdot10^{-4}$	$1.21\cdot10^{-8}$	$2.226\cdot10^{-3}$	0
11	$1.268\cdot10^{-8}$	0	0	$3.183\cdot10^{-15}$	$1.125\cdot10^{-4}$	0	1	0
12	$1.125\cdot10^{-4}$	0	0	0	0.998	$2.221\cdot10^{-3}$	$1.21\cdot10^{-8}$	0
13	$2.504\cdot10^{-7}$	$5.183\cdot10^{-6}$	0	0	$1.073\cdot10^{-4}$	$5.844\cdot10^{-10}$	1	0
14	0	0	0	0	0.048	$2.823\cdot10^{-11}$	0.952	0
15	$5.183\cdot10^{-6}$	0	0	0	0.046	$2.823\cdot10^{-11}$	0.954	0
16	$1.073\cdot10^{-4}$	$1.21\cdot10^{-8}$	$3.183\cdot10^{-15}$	0	0.046	0	0.954	0
17	$1.21\cdot10^{-8}$	0	$1.364\cdot10^{-12}$	0	0.046	$2.221\cdot10^{-3}$	0.952	0
18	$1.21\cdot10^{-8}$	0	0	0	$5.434\cdot10^{-6}$	$1.073\cdot10^{-4}$	0.048	0.952
19	$2.51\cdot10^{-7}$	0.046	0	0	0.952	0	$2.328\cdot10^{-3}$	0
20	$1.21\cdot10^{-8}$	0	0	0	0.048	0	0.952	0
21	$5.183\cdot10^{-6}$	0	0	0	0.046	0.952	$2.221\cdot10^{-3}$	0
22	0	0.046	$2.221\cdot10^{-3}$	0.952	$5.183\cdot10^{-6}$	$1.073\cdot10^{-4}$	$6.141\cdot10^{-10}$	0
23	$6.589\cdot10^{-14}$	0	$1.21\cdot10^{-8}$	0	0.048	0.952	$2.504\cdot10^{-7}$	0
24	0.046	$5.183\cdot10^{-6}$	0	0	$2.51\cdot10^{-7}$	0	0.954	0
25	$1.073\cdot10^{-4}$	0.046	0	0	$1.271\cdot10^{-8}$	$5.434\cdot10^{-6}$	0.954	0
26	$2.221\cdot10^{-3}$	0	0	0	$1.125\cdot10^{-4}$	$2.504\cdot10^{-7}$	0.998	0
27	0	0	0	0	0.046	$1.21\cdot10^{-8}$	0.954	0
28	0.046	0	$2.221\cdot10^{-3}$	0	$6.14\cdot10^{-10}$	0	0.952	0
29	0.952	0	$5.183\cdot10^{-6}$	0	0.046	$1.21\cdot10^{-8}$	$2.329\cdot10^{-3}$	0
30	0.046	0	0	0	$2.221\cdot10^{-3}$	$2.823\cdot10^{-11}$	0.952	0
31	$2.334\cdot10^{-3}$	$3.183\cdot10^{-15}$	0	0	0.952	0	0.046	0
32	$1.073\cdot10^{-4}$	0	0	0	0.046	0	0.954	0
33	0.046	0	0	0	$3.191\cdot10^{-15}$	$2.823\cdot10^{-11}$	0.954	0

A =

	1	2	3	4	5	6	7	8
33	0.046	0	0	0	$3.191\cdot10^{-15}$	$2.823\cdot10^{-11}$	0.954	0
34	0	$1.364\cdot10^{-12}$	0	0	$2.221\cdot10^{-3}$	$5.184\cdot10^{-6}$	0.998	0
35	$2.221\cdot10^{-3}$	0.046	0	$1.21\cdot10^{-8}$	$1.125\cdot10^{-4}$	$2.504\cdot10^{-7}$	0.952	0
36	$5.844\cdot10^{-10}$	$5.183\cdot10^{-6}$	0	0	0.046	0	0.954	0
37	$6.589\cdot10^{-14}$	0	0.998	0	$2.334\cdot10^{-3}$	$2.823\cdot10^{-11}$	$5.858\cdot10^{-10}$	0
38	$2.823\cdot10^{-11}$	0	$1.21\cdot10^{-8}$	0	0.952	$2.504\cdot10^{-7}$	0.048	0
39	$2.221\cdot10^{-3}$	0	0	0	0.998	$1.21\cdot10^{-8}$	$6.922\cdot10^{-14}$	0
40	0	$1.21\cdot10^{-8}$	0	0	$2.221\cdot10^{-3}$	$5.183\cdot10^{-6}$	0.998	0
41	$1.268\cdot10^{-8}$	0	0.952	0	$2.226\cdot10^{-3}$	0	0.046	0
42	0.048	0	0	0	$5.446\cdot10^{-6}$	0	0.952	0
43	$1.073\cdot10^{-4}$	$6.589\cdot10^{-14}$	$5.844\cdot10^{-10}$	0	$1.21\cdot10^{-8}$	0	1	0
44	$2.221\cdot10^{-3}$	$5.183\cdot10^{-6}$	0	0	0.998	0	$2.83\cdot10^{-11}$	0
45	0.046	0	0	0	0.954	0	$1.073\cdot10^{-4}$	0
46	0.952	$1.21\cdot10^{-8}$	0	0	$1.075\cdot10^{-4}$	0	0.048	0
47	$1.21\cdot10^{-8}$	0	0	0	$2.504\cdot10^{-7}$	$1.073\cdot10^{-4}$	1	$5.183\cdot10^{-6}$
48	0.046	$1.364\cdot10^{-12}$	$2.504\cdot10^{-7}$	0	0.954	0	$5.184\cdot10^{-6}$	$1.073\cdot10^{-4}$
49	$5.196\cdot10^{-6}$	0	0	0	$1.073\cdot10^{-4}$	$2.504\cdot10^{-7}$	1	0
50	$2.221\cdot10^{-3}$	0	0	0	0.998	0	$1.127\cdot10^{-4}$	0
51	0	0	0	0	$2.334\cdot10^{-3}$	0	0.998	0
52	0.046	$1.364\cdot10^{-12}$	0	0	$2.221\cdot10^{-3}$	$1.073\cdot10^{-4}$	0.952	0
53	0	0	0	0	0.048	$2.504\cdot10^{-7}$	0.952	0
54	0.952	$1.073\cdot10^{-4}$	0.046	0	$5.196\cdot10^{-6}$	0	$2.221\cdot10^{-3}$	0
55	$2.504\cdot10^{-7}$	$2.823\cdot10^{-11}$	0	0	0.954	0	0.046	$1.364\cdot10^{-12}$
56	$1.364\cdot10^{-12}$	0	0	$2.823\cdot10^{-11}$	0.952	$1.073\cdot10^{-4}$	0.048	0
57	$1.073\cdot10^{-4}$	0	0	0	0.048	0	0.952	$5.844\cdot10^{-10}$
58	0.952	0	$5.844\cdot10^{-10}$	0	0.048	0	$5.183\cdot10^{-6}$	0
59	0.998	0	0	0	$1.125\cdot10^{-4}$	$2.504\cdot10^{-7}$	$2.221\cdot10^{-3}$	0
60	0.954	0	0	0	$1.213\cdot10^{-8}$	$1.073\cdot10^{-4}$	0.046	0
61	0.952	0	0	0	0.046	$2.221\cdot10^{-3}$	$5.446\cdot10^{-6}$	0
62	0.998	0	0	$6.126\cdot10^{-10}$	$2.221\cdot10^{-3}$	$5.183\cdot10^{-6}$	$1.075\cdot10^{-4}$	0
63	0	$1.21\cdot10^{-8}$	0	0	0.954	$1.364\cdot10^{-12}$	0.046	0
64	$5.844\cdot10^{-10}$	0	0	0	0.998	$2.504\cdot10^{-7}$	$2.226\cdot10^{-3}$	0
65	$2.504\cdot10^{-7}$	0	0	0	0.952	$3.19\cdot10^{-15}$	0.048	$5.183\cdot10^{-6}$

	1	2	3	4	5	6	7	8
63	0	$1.21\cdot10^{-8}$	0	0	0.954	$1.364\cdot10^{-12}$	0.046	0
64	$5.844\cdot10^{-10}$	0	0	0	0.998	$2.504\cdot10^{-7}$	$2.226\cdot10^{-3}$	0
65	$2.504\cdot10^{-7}$	0	0	0	0.952	$3.19\cdot10^{-15}$	0.048	$5.183\cdot10^{-6}$
66	$2.221\cdot10^{-3}$	0	0	0	0.952	$2.823\cdot10^{-11}$	0.046	0
67	$5.844\cdot10^{-10}$	0	0	0	1	$1.364\cdot10^{-12}$	$2.625\cdot10^{-7}$	0
68	$2.504\cdot10^{-7}$	0	0	$5.183\cdot10^{-6}$	0.998	0	$2.328\cdot10^{-3}$	0
69	$1.21\cdot10^{-8}$	0	0	0	$2.51\cdot10^{-7}$	$2.221\cdot10^{-3}$	0.998	0
70	$1.21\cdot10^{-8}$	0	0	0	1	$1.073\cdot10^{-4}$	0	$3.183\cdot10^{-15}$
71	0	$5.183\cdot10^{-6}$	$6.604\cdot10^{-14}$	0	$5.844\cdot10^{-10}$	$2.221\cdot10^{-3}$	0.998	0
72	0	0	$2.823\cdot10^{-11}$	0	$6.589\cdot10^{-14}$	$3.183\cdot10^{-15}$	1	0
73	0.046	0	0	$5.844\cdot10^{-10}$	$1.364\cdot10^{-12}$	0	0.954	0
74	$1.364\cdot10^{-12}$	0	0	0	0.954	$1.268\cdot10^{-8}$	0.046	0
75	$2.221\cdot10^{-3}$	0.952	0	0	$1.367\cdot10^{-12}$	$1.073\cdot10^{-4}$	0.046	0
76	0	$1.364\cdot10^{-12}$	$1.21\cdot10^{-8}$	0	1	$2.504\cdot10^{-7}$	$1.125\cdot10^{-4}$	0
77	$1.073\cdot10^{-4}$	0.952	0	0	$2.226\cdot10^{-3}$	$2.823\cdot10^{-11}$	0.046	0
78	0	0	0	0	0.952	0.048	$2.96\cdot10^{-11}$	$3.183\cdot10^{-15}$
79	0	0.952	0	$6.589\cdot10^{-14}$	$2.504\cdot10^{-7}$	$2.221\cdot10^{-3}$	0.046	0
80	$2.221\cdot10^{-3}$	0	0	0	0.952	0	0.046	0
81	0.046	$3.183\cdot10^{-15}$	0	0	$2.625\cdot10^{-7}$	0.952	$2.334\cdot10^{-3}$	0
82	$5.844\cdot10^{-10}$	0	0	0	0.998	$1.073\cdot10^{-4}$	$2.221\cdot10^{-3}$	$6.589\cdot10^{-14}$
83	0	$1.21\cdot10^{-8}$	0.046	0	$2.328\cdot10^{-3}$	0.952	$5.434\cdot10^{-6}$	0

A =

Приложение 2. Программа расчета матрицы A.

```
Koeff_mult__Nn__(XNx, T, w) := | N ← rows(XNx)
                               | K ← cols(XNx)
                               | Koeff ← max(XNx^⟨K⟩)
                               | return "no length"  if length(w) ≠ 8
                               | for j ∈ 1..N
                               |   for i ∈ 1..Koeff
                               |     S_{j,i} ← 0
                               | n1 ← 0
                               | Q2 ← Σ_{q=1}^{2} w_q
                               | Q3 ← Σ_{q=1}^{3} w_q
                               | Q4 ← Σ_{q=1}^{4} w_q
                               | Q5 ← Σ_{q=1}^{5} w_q
                               | Q6 ← Σ_{q=1}^{6} w_q
                               | Q7 ← Σ_{q=1}^{7} w_q
                               | while (n1 < N)
                               |   | n1 ← n1 + 1
                               |   | for n ∈ 1..100
                               |   |   | V ← (1 + T)^(−n)
                               |   |   | t ← rnd(1)
                               |   |   | L ← 1  if 0 ≤ t < w_1
                               |   |   | L ← 2  if w_1 ≤ t < Q2
                               |   |   | L ← 3  if Q2 ≤ t < Q3
                               |   |   | L ← 4  if Q3 ≤ t < Q4
                               |   |   | L ← 5  if Q4 ≤ t < Q5
                               |   |   | L ← 6  if Q5 ≤ t < Q6
                               |   |   | L ← 7  if Q6 ≤ t < Q7
                               |   |   | L ← 8  otherwise
                               |   |   | S_{n1,L} ← S_{n1,L} + V
                               |   | Mak ← max[(S^T)^⟨n1⟩]
                               |   | Km ← XNx_{n1,K}
                               |   | if (S_{n1,Km} ≠ Mak)
                               |   |   | for i2 ∈ 1..Koeff
                               |   |   |   S_{n1,i2} ← 0
                               |   |   | n1 ← n1 − 1
                               | S ← S · T
                               | S
```

```
Koeff_mult__Nn__(XNx, T, w) := | N ← rows(XNx)
                               | K ← cols(XNx)
                               | Koeff ← max(XNx^⟨K⟩)
                               | return "no length"  if length(w) ≠ 8
                               | for j ∈ 1.. N
                               |   for i ∈ 1.. Koeff
                               |     S_(j,i) ← 0
                               | n1 ← 0
                               | Q2 ← Σ_(q=1)^2 w_q
                               | Q3 ← Σ_(q=1)^3 w_q
                               | Q4 ← Σ_(q=1)^4 w_q
                               | Q5 ← Σ_(q=1)^5 w_q
                               | Q6 ← Σ_(q=1)^6 w_q
                               | Q7 ← Σ_(q=1)^7 w_q
                               | while (n1 < N)
                               |   | n1 ← n1 + 1
                               |   | for n ∈ 1.. 100
                               |   |   | V ← (1 + T)^(−n)
                               |   |   | t ← rnd(1)
                               |   |   | L ← 1  if 0 ≤ t < w_1
                               |   |   | L ← 2  if w_1 ≤ t < Q2
                               |   |   | L ← 3  if Q2 ≤ t < Q3
                               |   |   | L ← 4  if Q3 ≤ t < Q4
                               |   |   | L ← 5  if Q4 ≤ t < Q5
                               |   |   | L ← 6  if Q5 ≤ t < Q6
                               |   |   | L ← 7  if Q6 ≤ t < Q7
                               |   |   | L ← 8  otherwise
                               |   |   | S_(n1,L) ← S_(n1,L) + V
                               |   | Mak ← max[(S^T)^⟨n1⟩]
                               |   | Km ← XNx_(n1,K)
                               |   | if (S_(n1,Km) ≠ Mak)
                               |   |   | for i2 ∈ 1.. Koeff
                               |   |   |   S_(n1,i2) ← 0
                               |   |   | n1 ← n1 − 1
                               | S ← S · T
                               | S
```

Приложение 3. Протофрактал Z.

$$Z = \begin{pmatrix} 10.276 & 3.313 & 2.59 & 94.907 & 27.899 & 0.065 & 0.035 & 6.253 \\ 20.638 & 1.688 & 1.289 & 200.858 & 34.995 & 0.046 & 0.006 & 8.455 \\ 37.262 & 0.442 & 0.164 & 173.095 & 71.949 & -0.003 & -0.001 & 45.915 \\ 9.955 & 1.709 & 1.191 & 39.852 & 96.355 & 0.389 & -0 & 9.738 \\ 21.437 & 1.667 & 1.335 & 100.208 & 26.923 & 0.023 & 0.002 & 10.876 \\ 12.33 & 2.296 & 1.827 & 96.794 & 35.277 & 0.279 & 0.007 & 6.367 \\ 15.02 & 2.23 & 1.883 & 109.989 & 26.779 & 0.029 & 0.007 & 9.177 \\ 9.695 & 4.28 & 3.953 & 91.895 & 10.636 & 0.125 & 0.211 & 2.321 \end{pmatrix}$$

Приложение 4. Сравнения средних значений и точек протофрактала.

$$\mathrm{Mean}(C1) = \begin{pmatrix} 10.58889 \\ 3.25444 \\ 2.54889 \\ 96.26333 \\ 27.8 \\ 0.06444 \\ 0.03333 \\ 6.45556 \end{pmatrix} \qquad \left(Z^T\right)^{\langle 1 \rangle} = \begin{pmatrix} 10.276 \\ 3.313 \\ 2.59 \\ 94.907 \\ 27.899 \\ 0.065 \\ 0.035 \\ 6.253 \end{pmatrix}$$

$$\mathrm{Mean}(C2) = \begin{pmatrix} 20.7 \\ 1.7 \\ 1.3 \\ 196.9 \\ 35.5 \\ 0 \\ 0 \\ 9.1 \end{pmatrix} \qquad \left(Z^T\right)^{\langle 2 \rangle} = \begin{pmatrix} 20.638 \\ 1.688 \\ 1.289 \\ 200.858 \\ 34.995 \\ 0.046 \\ 6.459\times 10^{-3} \\ 8.455 \end{pmatrix}$$

$$\mathrm{Mean}(C3) = \begin{pmatrix} 36.1 \\ 0.55 \\ 0.26 \\ 169.95 \\ 70.05 \\ 0 \\ 0 \\ 43.95 \end{pmatrix} \qquad \left(Z^T\right)^{\langle 3 \rangle} = \begin{pmatrix} 37.262 \\ 0.442 \\ 0.164 \\ 173.095 \\ 71.949 \\ -2.507\times 10^{-3} \\ -8.093\times 10^{-4} \\ 45.915 \end{pmatrix}$$

$$\mathrm{Mean}(C4) = \begin{pmatrix} 10.2 \\ 1.74 \\ 1.23 \\ 43.44 \\ 92.7 \\ 0.37 \\ 0 \\ 9.7 \end{pmatrix} \qquad \left(Z^T\right)^{\langle 4 \rangle} = \begin{pmatrix} 9.955 \\ 1.709 \\ 1.191 \\ 39.852 \\ 96.355 \\ 0.389 \\ -4.029\times 10^{-4} \\ 9.738 \end{pmatrix}$$

$$\mathrm{Mean}(C5) = \begin{pmatrix} 21.189 \\ 1.691 \\ 1.356 \\ 100.568 \\ 27.074 \\ 0.024 \\ 2.222\times 10^{-3} \\ 10.811 \end{pmatrix} \qquad \left(Z^{T}\right)^{\langle 5\rangle} = \begin{pmatrix} 21.437 \\ 1.667 \\ 1.335 \\ 100.208 \\ 26.923 \\ 0.023 \\ 1.787\times 10^{-3} \\ 10.876 \end{pmatrix}$$

$$\mathrm{Mean}(C6) = \begin{pmatrix} 12.48571 \\ 2.29286 \\ 1.82429 \\ 96.8 \\ 34.68571 \\ 0.26714 \\ 0.00714 \\ 6.58571 \end{pmatrix} \qquad \left(Z^{T}\right)^{\langle 6\rangle} = \begin{pmatrix} 12.33 \\ 2.296 \\ 1.827 \\ 96.794 \\ 35.277 \\ 0.279 \\ 6.849\times 10^{-3} \\ 6.367 \end{pmatrix}$$

$$\mathrm{Mean}(C7) = \begin{pmatrix} 15.1 \\ 2.2 \\ 1.9 \\ 109.7 \\ 26.9 \\ 0 \\ 0 \\ 9.2 \end{pmatrix} \qquad \left(Z^{T}\right)^{\langle 7\rangle} = \begin{pmatrix} 15.02 \\ 2.23 \\ 1.883 \\ 109.989 \\ 26.779 \\ 0.029 \\ 6.524\times 10^{-3} \\ 9.177 \end{pmatrix}$$

$$\mathrm{Mean}(C8) = \begin{pmatrix} 10 \\ 4.17 \\ 3.84 \\ 92.84 \\ 11.5 \\ 0.12 \\ 0.2 \\ 2.7 \end{pmatrix} \qquad \left(Z^{T}\right)^{\langle 8\rangle} = \begin{pmatrix} 9.695 \\ 4.28 \\ 3.953 \\ 91.895 \\ 10.636 \\ 0.125 \\ 0.211 \\ 2.321 \end{pmatrix}$$