

АЗЕРБАЙДЖАНСКАЯ РЕСПУБЛИКА

На правах рукописи

ПРИМЕНЕНИЕ ХАОТИЧЕСКИХ ВРЕМЕННЫХ РЯДО ДЛЯ ПРОГНОЗИРОВАНИЯ ПОТРЕБИТЕЛЬСКОГО СПРОСА

Специальность: 3338.01 – “Системный анализ, управление и обработка информации” (управление и принятие решений)

Отрасль науки: Технические науки

Соискатель: **Мустафаева Севиндж Расул кызы**

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора философии

Баку – 2022

Диссертационная работа выполнена на кафедре «Компьютерная инженерия» Азербайджанского Государственного Университета Нефти и Промышленности.

Научный руководитель: доктор технических наук, профессор
Джавадов Натик Гаджи оглы

Официальные
оппоненты:

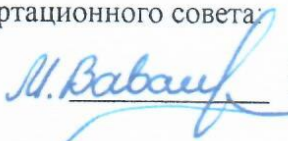
доктор технических наук, профессор
Нуриев Мухаммед Нурмухаммед оглы

доктор технических наук, профессор
Юсифов Салахаддин Имамали оглы

доктор технических наук, профессор
Мамедов Джаваншир Фируддин оглы

Диссертационный совет ED 2.02 Высшей Аттестационной Комиссии при Президенте Азербайджанской Республики, действующей на базе Азербайджанского Государственного Университета Нефти и Промышленности.

Председатель
диссертационного совета:



доктор технических наук, профессор
Бабанлы Мустафа Баба оглы

Ученый секретарь
диссертационного совета:



кандидат технических наук, доцент
Джаббаров Тахир Гафар оглы

Председатель научного
семинара



доктор технических наук, профессор
Абдуллаев Тарлан Самед оглы



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Проблема анализа и обработки временных рядов в контексте принципов синергетики приобретает особую актуальность.

Временные ряды являются основным результатом экспериментов как натуральных, так и вычислительных. В качестве приоритетных обычно указывают задачи метеорологии, геофизики, финансового анализа. В последнее время присоединились физиология, медицина и социальные науки. При обработке временных рядов обычно решают несколько типов задач. В настоящее время продолжает оставаться актуальной проблема анализа и прогнозирования большого количества гетерогенных взаимодействующих информационных потоков в сложных сетевых структурах. Эволюционные траектории в пространстве состояний системы становятся чувствительными к малым информационным воздействиям (флуктуациям). В результате кумулятивного эффекта в определенный момент система может переходить с одной траектории эволюционного развития на другую, причем ее поведение определяется не энергией воздействий, а их параметрами и формой.

В процессе эволюции в открытой системе рост информационных потоков приводит к усложнению информационной составляющей, что в свою очередь вызывает нарастание хаотических процессов, переводящих систему в состояние динамического хаоса. Кроме того, гетерогенная составляющая информационных потоков оказывает значительное влияние на процессы когерентности (в широком смысле), что снижает возможности эффективного функционирования системы.

В прикладном анализе наблюдаемых выделяются следующие основные задачи: получение наблюдаемой, задача идентификации и прогнозирования.

Поскольку прогнозирование представляет широкий спектр мыслей и мнений - как и рынки в целом, рынки предсказаний оказались весьма эффективными в качестве прогностического инструмента. В результате их дальновидности рынки прогнозов

(иногда называемые виртуальными рынками) использовались рядом крупных компаний. Смещение экономики, политики и, в последнее время, культурных факторов только увеличило спрос на прогноз. Добавьте преимущества аналитики данных и искусственного интеллекта; мы живем в золотой век данных и статистической полезности.

Однако традиционные методы реализации этих задач требуют значительного объема информации либо стационарных рядов данных, которые не всегда можно получить на практике при исследовании реальных систем. Отсюда проблема идентификации и прогнозирования наблюдаемых, в контексте принципов синергетики, приобретает особую актуальность.

Объект и предмет исследования. Определение потребительского спроса автомобильного объекта на основе нечеткой логики в условиях неопределенности, анализ связи между фрактальностью и нечеткой модели.

Цель и задачи работы. Целью диссертации является прогнозирование потребительского спроса на основе хаотических временных данных в автомобильной области опираясь на данные продаж предыдущих лет.

Для достижения поставленных автором целей были поставлены и решены следующие задачи:

- представление данных по продажам автомобилей в виде временных рядов;
- определение положительного показателя Ляпунова временных рядов, описывающих продажи автомобилей;
- определение состояния хаотичности данных временных рядов;
- фрактальный анализ хаотических временных рядов и определение фрактальной размерности;
- анализ эффекта долговременной памяти хаотических временных рядов;
- определение связи между фрактальной размерностью и количеством правил моделей нечеткой логики;
- прогнозирование потребительского спроса автомобильных продаж с использованием моделей ЕСЛИ-ТО.

Методы исследований. В качестве методов исследования в диссертации использовались синергетические принципы, нелинейную динамику, теорию хаоса; теорию управления хаотическими процессами, функциональный анализ; методы фрактального анализа, методы нечеткой логики.

Проведено компьютерное моделирование в среде MATLAB и подтверждена точность полученных результатов.

• **Основные положения, выносимые на защиту.** В диссертации были рассмотрены следующие вопросы:

- разработка математической модели системы, учитывающая новое прочтение хаоса.

- разработка алгоритма непараметрического прогнозирования на основе вычисления фрактальной размерности и возможностей нечеткой логики.

- проведение кластерного анализа хаотических временных рядов с учетом связи фрактальности с количеством правил;

- с учетом правила “Если-То” нечеткой модели прогнозирование спроса с использованием хаотических временных рядов.

Научная новизна. Основными научными инновациями, полученными в диссертации, являются:

- Проведено многомерное исследование по идентификации временных рядов по продажам нескольких лет.

- Проведен анализ динамического поведения системы с учетом нового прочтения теории хаоса.

- Определена фрактальность временных хаотических рядов и доказан эффект долговременной памяти.

- Установлена связь между фрактальной размерностью и количеством правил нечеткой модели.

- Впервые изучена нечеткая устойчивость финансовых временных рядов, характеризующаяся неопределенностью.

- С учетом правила “Если-То” нечеткой модели спрогнозирован потребительский спрос финансовых хаотических временных рядов.

Научно-практическая значимость. Теоретическая значимость заключается в том, что разработка системы «наблюдаемые-идентификация-управление» и включение инвариантных множеств в синергетические методы анализа и управления хаотическими и системами вносят весомый вклад, дополняющий существующие традиционные методы анализа наблюдаемых в пространстве управления.

Использование метода нечеткой логики позволяет не предъявлять особых требований к данным и позволяет визуально исследовать поведение наблюдаемых и сделать прогноз.

Научные результаты, полученные в диссертации, могут быть применены к маркетингу в различных сферах. Основная практическая значимость заключается в возможности определения спроса, что в свою очередь позволяет планирование бюджета и снижение затрат.

Апробация работы. Результаты, полученные в диссертационной работе, были заслушаны на следующих конференциях:

1. Мустафаева С.Р. Графические методы идентификации и прогнозирования взаимодействующих временных последовательностей // Международная научно – практическая конференция молодых ученых и студентов «Информационные процессы и технологии, - Севастополь: Изд-во СевНТУ,- 23-27 апреля, - 2012, стр.267.

2. Мустафаева С.Р. Анализ взаимодействующих хаотических временных рядов //Третья Международная научно-практическая конференция «Информационные технологии и компьютерная инженерия», - Винница: Изд-во ВНТУ 29-31 мая 2012 стр.262-263.

3. Мустафаева С.Р. Anomalous diffusion influence of chaotic system by observation // IV International conference “Problems of cybernetics and informatics”,- Baku,- September 12-14 ,2012, pp.158-160.

4. Мустафаева С.Р. Идентификация и прогнозирование временных рядов // Doktorantların və gənc tədqiqatçıların Azərbaycan xalqının Ümummilli lideri Heydər Əliyevin anadan olmasının 90 iliyinə

həsr olunmuş” Azərbaycan 2020: neft-qaz sənayesinin inkişaf perspektivləri” konfrans ,- Bakı: - 2-3 may ,- 2013, s.183-185.

5. Мустафаева С.Р. Нелинейные методы обработки временных рядов //Международная научно-практическая конференция молодых ученых, аспирантов и студентов, - Харьков: Изд-во, - 2013, стр.82.

6. Mustafayeva S.R. Calculation of Lyapunov exponent of time series //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-2019 Switzerland, pp.768-771.

7. Mustafayeva S.R. Fractal Dimension Determining for Demand Forecasting //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW- Switzerland , pp.931-937.

8. Мустафаева С.Р. Использование метода кластеризации в автомобильном бизнесе, ADNSU-nun 100 illiyinə həsr olunmuş ADNSU gənc alim və tədqiqatçıların elmi-praktiki konfransı 2020-Bakı: - 2020 ADNSU, стр.452-457.

Публикации. По теме диссертации опубликовано 18 научных работ: 8 статей опубликовано в рецензируемых научных журналах, 10 опубликовано в сборниках материалов конференций. Основная индексация перечисленных работ представлена следующим образом:

2 статьи опубликованы в сборнике материалов конференции, индексируемом в Web of Science.

2 статьи были напечатаны в журналах, рекомендованы АР ААК.

6 статей были напечатаны в журналах включенных в РИНЦ и рекомендованных ВАК РФ.

Название организации, в которой выполнена диссертационная работа. Диссертационная работа выполнена на кафедре «Компьютерная инженерия» Азербайджанского Государственного Университета Нефти и Промышленности.

Структура и объем диссертации . Диссертация состоит из введения, 4 глав, заключения, списка из 150 использованных

источников и приложений. Работа написана на 155 страницах, в том числе 3 таблицы и 13 рисунков. Общий объем диссертации составляет 175 000 знаков.

СОДЕРЖАНИЕ РАБОТЫ

Во введении представлены актуальность работы, научная новизна, методы исследований, основные защищаемые положения и практическая значимость диссертации.

В первой главе даны состояние, анализ проблем и возможностей временных рядов при реализации различных задач маркетинга и программы их действий. Сформулированы цель и задачи, необходимые для реализации работы.

Продажа - жизненно важная часть любого бизнеса. Соответственно прогнозирование продажи играет важную роль в финансах бизнеса планирования и является инструментом самооценки компании. Менеджеры должны держать руку на пульсе своей компании. Отчеты с прогнозом продаж, графики анализирует пульс бизнеса. Это может иметь значение между просто выживанием и высоким успехом в бизнес. Это важный краеугольный камень бюджета компании. В будущем, направление компании может зависеть от точности прогноз продаж. Прогнозирование позволяет компаниям эффективно распределять ресурсы для будущего роста и управлять денежными потоками.

Один из первых достижений в этой области это работа Уайта (1988), а также другие соответствующие статьи, в которых используются сложные сети: Bosarge (1993), Wong (1990), Niemstra (1996) и Хефке и Helmenstein (1996), которые находят нелинейности во временных рядах и улучшают прогнозы. Другим вкладом с прогнозами для других макроэкономических переменных являются работы Franses and Draisma (1997) и Swanson and White (1997). Прогнозирование может быть выполнено для закрытия/открытия курса акций на ежедневной основе, квартальных доходов компании и т.д. В литературе доступны различные модели для прогнозирования данных временных рядов. Некоторые из них:

- скользящее среднее авторегрессионной интеграции (ARIMA);
- простая скользящая средняя (SMA);
- экспоненциальное сглаживание (SES);
- нейронная сеть (NN);
- модели линейной регрессии;
- логистическая регрессия;
- машина опорных векторов;
- наивный байесовский;
- скрытый Марков;
- VAR;
- Гауссовы процессы.

Требуются методы, которые относительно просто реализовать и легко интерпретировать. В идеале результаты предварительного анализа должны указывать направления дальнейшего анализа, а оттуда - дальнейшие направления исследований. Синтез новых методов анализ в последовательной, унифицированной форме позволит исследователям проводить анализ более качественно и уверенно. Подход временных рядов наиболее эффективен для устоявшихся предприятий, располагающих данными за несколько лет, на основе которых можно работать наряду с относительно стабильными моделями тенденций.

Исходя из вышеизложенного анализа в области продаж автомобилей можем еще раз подчеркнуть необходимость исследований в прогнозировании потребительского спроса. Модель, используемая в этой работе, не требует специальных знаний о используемом наборе данных. Преимущество использования нечеткой логики заключается в обработке неточных, неопределенных, нечетких, полу истинных или приближенных и нелинейных данных.

С учетом отмеченных выше недостатков в диссертации сформулирована постановка задачи.

Целью исследования является прогнозирование потребительского спроса с помощью хаотических временных

рядов на основе нечеткой логики. Для достижения этой цели необходимо решить следующие проблемы:

- разработать математическая модель системы, учитывающая новое прочтение хаоса.
- разработать алгоритм непараметрического прогнозирования на основе вычисления фрактальной размерности и возможностей нечеткой логики.
- провести кластерный анализ хаотических временных рядов с учетом связи фрактальности с количеством правил;
- с учетом правила “Если-То” нечеткой модели спрогнозирован спрос с использованием хаотических временных рядов.

Вторая глава посвящена разработке интеллектуальной адаптивной системы «наблюдаемые-идентификация-управление».

В общем случае задача моделирования по временным рядам известна под именами «идентификация систем» и «прогноза систем» .

В работе состояние и перспективы реализации исследуемой проблемы будут рассматриваться в контексте задач идентификации и задач прогноза.

Задача идентификации. При ее решении делается попытка ответить на вопрос, каковы параметры системы, породившей данный временной ряд. Параметры могут быть самыми различными – статистические распределения, параметры статистических моделей, спектральные свойства и т.п. Важно, чтобы эти параметры помогли бы идентифицировать систему (процесс), то есть отличить ее от других.

Задача прогноза. Она состоит в том, чтобы по данным наблюдений предсказать будущие значения измеряемых характеристик или, более широко, будущее состояние анализируемого объекта.

Если малые возмущения начального условия с течением времени не нарастают (то есть имеет место устойчивость), то поведение такой системы предсказуемо. В противном случае процесс может быть описан только вероятностным образом. Именно эти соображения легли в основу современного представления о динамическом хаосе.

Важность изучения хаоса заключается в том, что хаос предлагает альтернативный метод, объясняющий очевидно случайное поведение сложных систем. Хаос плюс специфические математические инструменты - это основа для изучения различных моделей из разных областей, которые можно свести к элементарным с известным хаотическим поведением для некоторых значений параметров.

В общем случае скалярным временным рядом (ВР) будет называться массив из N чисел, представляющих собой значения некоторой динамической переменной $x(t)$ с постоянным шагом Δt по времени, то есть в моменты

$$t_i = t_0 + (i - 1)\Delta t; x_i = x(t_i), i = \overline{1, N}. \quad (1)$$

Временные ряды – это наборы данных, представляющие поведение одной (или нескольких) случайных величин во времени, и их основная характеристика состоит в том, что последовательные записи этой переменной не являются независимыми друг от друга, и их анализ должен учитывать порядок, в котором они были собраны.

Вычисление ляпуновских показателей по реализации. Одной из особенностей хаотических режимов является неустойчивость каждой траектории, принадлежащей хаотическому аттрактору.

Количественной мерой этой неустойчивости оказались характеристические показатели Ляпунова. Они являются одной из важнейших характеристик аттрактора, поскольку позволяют оценить:

- фрактальную размерность аттрактора;
- энтропию динамической системы;
- характерное время предсказуемости поведения системы.

В процессе эволюции в открытой системе рост взаимодействующих информационных потоков и объектов приводит к усложнению информационной составляющей, что в свою очередь вызывает нарастание хаотических процессов, которые переводят систему в состояние динамического хаоса

Поскольку критерием хаотической динамики служит присутствие положительного старшего ляпуновского показателя,

представляет большой интерес возможность его оценки на основании обработки временных рядов (реализаций). Пусть $x(t)$ - типичная фазовая траектория системы $\dot{x}=v(x,a)$ и $x_i(t)$ - близкая к ней траектория, $x_i(t) = x(t) + \xi(t)$ [14, стр.419-431] . Рассмотрим функцию

$$\Xi(\xi(0)) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{|\xi(t)|}{|\xi(0)|}, \quad (2)$$

которая определена на векторах начального смещения $\xi(0)$ таких, что $|\xi(0)| = \varepsilon$, где $\varepsilon \rightarrow 0$. Тогда в зависимости от направления вектора $\xi(0)$ функция $\Xi(\xi(0))$ будет принимать конечный ряд значений $\{\lambda_i\}$, $i = 1, 2, \dots, n$, которые называются *характеристическими показателями Ляпунова* .

Характеристические показатели Ляпунова служат мерой хаотичности динамических систем. В частности, если имеются положительные показатели, то поведение системы будет хаотическим.

Рассмотрим временной ряд, полученный в результате записи наблюдаемой переменной величины в последовательные равноотстоящие моменты времени x_1, x_2, \dots, x_N .

Зададимся некоторым целым m и используем идею, реконструкции аттрактора методом запаздывания в пространстве размерности m , то есть положим

$$x_i = (x_1, x_{i-p}, x_{i-2p}, \dots, x_{i-(m-1)p}), \quad i = 1, 2, \dots \quad (3)$$

Рассматривая множество точек x_i в m - мерном пространстве как фазовый портрет аттрактора, оценим его размерность по методу Грасбергера-Прокачиа ,то есть, вычислим для разных ε корреляционный интеграл

$$\tilde{N}(\varepsilon) = \lim_{M \rightarrow \infty} \frac{1}{M(M-1)} \sum_{i,j=1}^M \theta(\varepsilon - \|x_i - x_j\|) \quad (4)$$

отложим полученную зависимость на графике в координатах $(\log \varepsilon, \log C(\varepsilon))$, произведем поиск на ней линейного участка и определим угловой коэффициент $D(m)$. Описанную процедуру

выполним несколько раз, рассматривая последовательно $m = 1, 2, 3, \dots$

Здесь важно отметить, что наличие или отсутствие насыщения зависимости $D(m)$ при увеличении m рассматривается как критерий того, генерируется ли сигнал динамической системой или же он является шумовым. Если наблюдается насыщение на некотором уровне D , то это величина D принимается в качестве оценки корреляционной размерности аттрактора ДС, породившей наблюдаемый сигнал. Основываясь на оценке D и теореме о вложении Мане, можно заключить, что размерность фазового пространства этой ДС не превышает $2D + 1$.

Утверждение, доказанное Мане, состоит в том, что любое практически гладкое отображение X в Y будет задавать вложение множества A в пространство Y при условии, что

$$m \geq 2D_A + 1, \quad (5)$$

где D_A - фрактальная размерность множества A ; m - размерность вложения.

Нечеткая логика основана на теории нечетких множеств, которая является обобщением классического множества теория (Заде, 1965). Проблема идентификации нечетких систем – проблема извлечения правил IF-THEN из сырых входных и выходных данных. Это происходит следующим образом:

- 1) Кластеризация
- 2) Спецификация отношений ввода-вывода (правила IF-THEN)

Третья глава посвящена интеллектуальному анализу и идентификации характеристик хаотических временных рядов описывающих динамику продаж автомобилей официальным дилером.

Как говорилось ранее, учитывая открытость и динамичность финансовых процессов, система описывается временными рядами. Данные представляют собой продажи автомобилей официальным дилером за 2012-2013 года.

Таблица 1
Данные по продажам автомобилей официальным дилером за два года

2012 год	2013 год
56	46
54	65
67	75
111	70
99	78
63	80
133	88
79	91
87	93
71	49
86	65
135	63

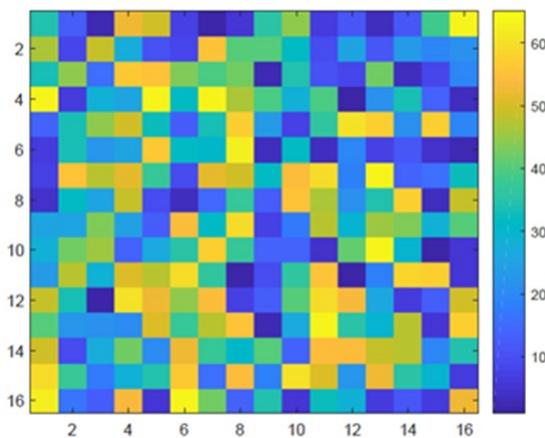


Рисунок 1. Фрагмент хаотических временных рядов по продажам автомобилей за 2 года

Для того чтобы доказать хаотичность наших временных рядов определяем показатель Ляпунова:

1. Вводим данные для продаж 2-х лет.
2. Используя программу в МАТЛАБ находим значение показателя Ляпунова

```
function zout = zlap(I,J,X,TAU)
% m file for function zlap used in lap.m for calculation of
% the first Lyapunov exponent using the method described by
Wolf
zout = X(I+(J-1)*TAU);
```

Lyapunov exponent
Lamda1=0.6609

Используя программный пакет МАТЛАБ определяем показатель Ляпунова $L = 0.6609$.

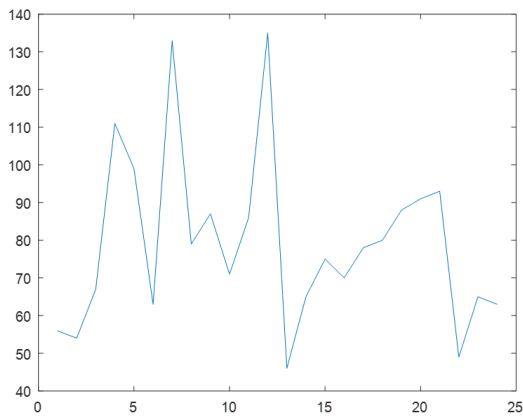


Рисунок 2. Графическое представление показателя Ляпунова

Положительный показатель Ляпунова показывает существование хаоса в динамической системе с ограниченными траекториями. Далее по алгоритму мы должны определить фрактальную размерность данной системы. Для начала необходимо раскрыть понятие фрактального анализа временных рядов.

Прежде чем перейти к определению фрактальной размерности введем понятие “фрактала“. Фрактал можно определить как ниже:

«Кривая или геометрическая фигура, каждая часть которой имеет тот же статистический характер, что и целое. Фракталы полезны при моделировании структур (таких как размытые береговые линии или снежинки), в которых похожие модели повторяются в постепенно уменьшающихся масштабах, а также при описании частично случайных или хаотических явлений, таких как рост кристаллов, турбулентность жидкости и образование галактик».

Существует множество методов количественной оценки фрактальных характеристик структуры через фрактальную размерность. В качестве традиционного примера фрактальная размерность пространственной фрактальной структуры может быть определена количественно с помощью фрактального анализа с подсчетом ящиков, который исследует способ, которым структура заполняет пространство. Однако такой пространственный анализ, как правило, не очень хорошо подходит для анализа так называемых фракталов «временных рядов», которые могут демонстрировать точную или статистическую самоаффинность, но которым по своей сути не хватает четко определенных пространственных характеристик.

Фрактал – это геометрический объект с внутренней гомотетикой (от греч. *ὁμοιο-θητικός*, подобная форма), что означает, что он повторяет его форму в той же форме в разных масштабах. Увеличение любой части, можно получить цифру, аналогичную оригиналу. Поэтому ее называют фрактальной геометрией. Фрактальная геометрия сводится к идентификации этих конфигураций, анализу и манипуляции ими. Также можно использоваться не только как инструмент для анализа и синтеза, но и как прогноз. Правила фракталов точны, поэтому результат предсказуем. Это контрастирует с традиционной наукой, которая вместо этого рассматривает нерегулярные аспекты природы и не похожие события.

Хорошо известно, что определение фрактальных размерностей систем дает возможность прогнозировать их поведение. В этой связи в работе было проведено исследование в контексте выбора области интересов, анализа и определения ее фрактальной размерности. В данной случае мы хотим определить фрактальную размерность временных рядов для прогнозирования спроса на рынке в следующем году по имеющимся данным за последние годы. В частности, такая характеристика временного ряда, как фрактальная размерность, позволяет определить момент, когда система становится нестабильной и готовой к переходу в новое состояние.

С помощью MATLAB Simulink определяется фрактальную размерность временных рядов. Сначала мы конвертируем наши данные из линейной формы в трехмерную. Затем определяем фрактальную размерность. Следовательно, временной ряд имеет эффект долговременной памяти, оценив его глубину, выявляем наличие циклов (квазициклов). И мы можем сказать, что наш результат характеризует вложения в соответствующий финансовый инструмент как «относительно без риска».

Поскольку это измерение говорит о хаосе временных рядов, мы можем определить фрактальную размерность $D=2.81$. Используя MATLAB Simulink, мы получаем фрактальную размерность временных рядов. Сначала мы конвертируем наши данные из линейной формы в трехмерную. Затем определяем фрактальную размерность. Следовательно, временной ряд имеет эффект долговременной памяти, оценив его глубину, выявляем наличие циклов (квазициклов). И мы можем сказать, что наш результат характеризует вложения в соответствующий финансовый инструмент как «относительно без риска».

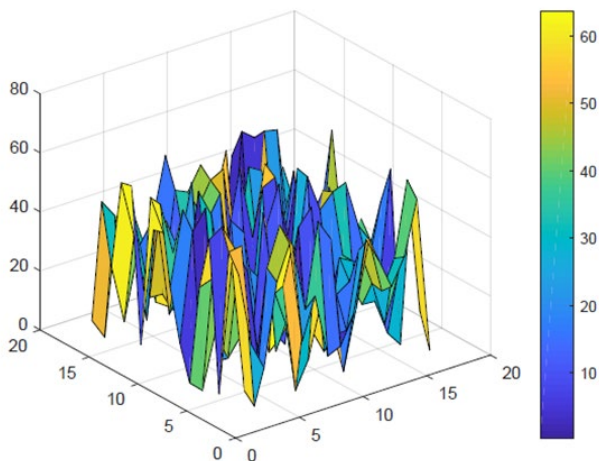


Рисунок 3. Фрактальная размерность полученная в MATLAB

Традиционно мы присваиваем конкретному геометрическому объекту четкое значение фрактала и это числовое значение считалось специфической меткой для объекта. Фрактальный анализ оказался полезным при описании динамических явлений. Это также привело к появлению ряда приложений. Сам термин относится к характеристике глобального свойства самоподобия, которое описывает, насколько часть явления может быть увеличена до целого. Инструмент, используемый для количественной оценки этого эффекта, - фрактальная размерность. Во временных рядах, например, фрактальная размерность выражает регулярность ряда и указывает, насколько сходство увеличивается, когда такие временные ряды наблюдаются в течение более длительного временного интервала. Самоподобие можно также рассматривать как меру геометрической сложности обсуждаемого объекта. Определение фрактальной размерности неразрывно связано с конструкциями, основанными на множестве.

Общим для существующих методов (несмотря на очевидные технические различия) является то, что все они используют

наборы, рассматриваемые как информационные гранулы, которые позволяют нам видеть только определенную часть явления. Изменение размера информационных гранул означает, насколько большую часть мы принимаем во внимание.

Гранулярные вычисления являются краеугольным камнем обработки в интеллектуальных системах. Человек воспринимает слово, систематизирует знания и делает их высокопроизводительными, формируя информационные гранулы и манипулируя ими. Информационная грануляция – это пример абстракции. Существует множество аспектов гранулярной обработки информации, а также существует множество формальных структур, в которых такое гранулирование информации происходит. Например, они включают теорию множеств, нечеткие множества, случайные множества, грубые множества и многие другие.

Становится совершенно очевидным, что могут существовать некоторые интересные и потенциально полезные связи между фундаментальными концепциями фракталов и гранулярными вычислениями. Фракталы используют информационные гранулы для определения своих основных характеристик.

В данном исследовании наша цель - выявить связи между фрактальным анализом и ролью информационных гранул, которые в нем играют. В частности, мы пересматриваем основную идею вычисления фрактальной размерности с использованием нечетких множеств.

В этом случае обычно используют корреляционное измерение, где мы строим гиперсферы вокруг отдельных точек набора данных, см. рисунок 6. Следуя этому пути, временная сложность зависит от количества элементов данных. Следуя этому методу, мы подсчитываем количество точек данных, находящихся в сфере радиуса s .

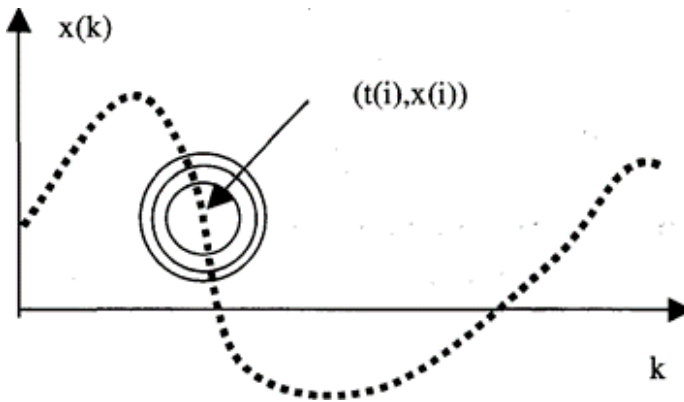


Рисунок 4. Вычисление корреляционной фрактальной размерности

В терминах фрактальной размерности измеряется сложность объекта, который может быть от природы (например, бактерии) или математический объект, созданный с помощью формулы или алгоритма (например, хорошо известного как Множество Мандельброта).

Однако такие числовые значения трудно использовать при классификации или приложений распознавания, поскольку рассчитанные значения не соответствуют этим четким значениям.

Исходя из этого, была предложена следующая схема для оценки нечеткой фрактальной размерности множества геометрических объектов. Во-первых, стандартная фрактальная размерность объектов рассчитывается с использованием алгоритма подсчета ящиков с выборками точки от объектов. Во-вторых, четкие значения фрактальных размерностей объектов используется для построения лингвистических значений для измерений, а именно нечетких фрактальных размерностей объектов. В качестве примера рассмотрим гауссову функцию принадлежности:

$$A_{ij}(e) = \exp\left[-(x(i) - x(f))^2 / e^2\right] \quad (6)$$

Предполагая, что $x(i)$ фиксировано, функция принадлежности возвращает степень принадлежности $x(j)$ этой окрестности. В этом

смысле он помогает различать (дискриминировать) различные элементы временного ряда, попадающие в данное окно $Q_{ij}(s)$. Чем больше степень принадлежности A_{tj} для данного $x(j)$, тем более заметен его вклад в общую сумму $N(s)$. Сама функция принадлежности оснащена параметром распространения (ϵ), который управляет размером информационной гранулы. Следует подчеркнуть, что нечеткие множества в приведенном выше определении задаются в амплитудном пространстве, при этом переменная времени, по которой распределены все дискретные моменты времени, не затрагивается, а грануляция там связана с окном (рисунок 5)

Далее, используя эти лингвистические значения фрактальных измерений, строится набор нечетких правил. так что каждый объект связан с каждым правилом. Таким образом, этот набор нечетких правил «если – то» можно рассматривать в качестве схемы классификации для набора геометрических объектов и может использовать для распознавания этих объектов, потому что конкретное значение отображается на объект. Это и есть связь между фрактальной размерностью и количеством информационных гранул, необходимых для использования в деятельности по моделированию.

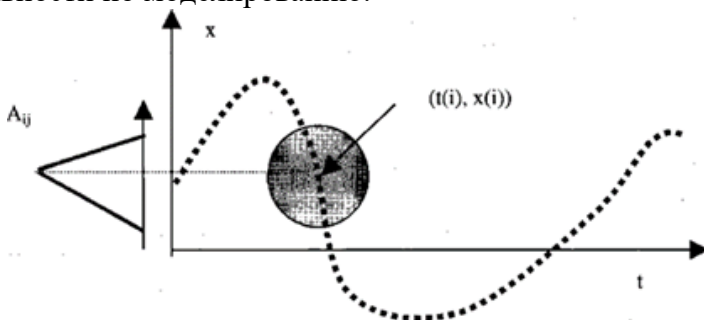


Рисунок 5. Вычисление фрактальной размерности с использованием корреляционной размерности в рамках нечетких множеств, используемых в качестве информационных гранул

Интуитивно понятно, что чем сложнее моделируемое явление, тем больше информационных гранул необходимо использовать для построения модели. Конъюнктура, которую мы делаем, заключается в том, что фрактальная размерность определяет количество нечетких множеств, а соотношение является степенным законом.

Вычисления локальной фрактальной размерности помогают нам идентифицировать области в пространстве данных, где нам нужны более «подробные» информационные гранулы, и, следовательно, количество правил, необходимых для моделирования данных там, становится больше. Степенной закон детализация информации становится полезным при распределении количества нечетких наборов по областям данных.

Четвертая глава посвящена компьютерной симуляции на основе кластерного анализа (c-means) для определения потребительского спроса в автомобильном бизнесе. Основным результатом работы является вывод о возможности проведения прогнозного анализа финансовых временных рядов путем расчета фрактальной меры. В то же время этот анализ не является исчерпывающим инструментом прогнозного исследования хаотических временных рядов, поскольку он не всегда дает полную информацию об их поведении без использования дополнительных методов и алгоритмов. Поэтому как и говорилось ранее далее используем возможности нечеткой логики в среде МАТЛАБ.

Кластерный анализ используется для 4 основных целей. Эти цели заключаются в следующем:

а) Разделите n объектов на различные подгруппы в максимально возможной степени, каждый со своими особенностями, определенными переменной r ;

б) разделить переменную r на подгруппы, которые, как считается, имеют общие черты в терминах значений, определенных в единицах n , и выявить структуры общих факторов;

в) Отдельные подмножества единицы, r , вместе с единицами и переменными вместе;

г) Определите биологическую и типологическую классификацию, которой следуют единицы для значений, определенных.

Этапы кластерного анализа:

а) Определение размеров и переменных, используемых для определения сходства между единицами;

б) Умножение единиц после нахождения сходства между единицами;

в) определение совместимости созданных кластеров;

г) статистическая достоверность воображения большинства населения

Этот алгоритм работает, назначая членство каждой точке данных, соответствующей каждому центру кластера, на основе расстояния между центром кластера и точкой данных. Чем ближе данные к центру кластера, тем больше их принадлежность к конкретному центру кластера. Ясно, что сумма членства каждой точки данных должна быть равна единице. После каждой итерации центры членства и кластеров обновляются по формуле:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{(2lm-1)}} \quad (7)$$

$$v_{ij} = \frac{\left(\sum_{i=1}^n (\mu_{ij})^m x_i \right)}{\left(\sum_{i=1}^n (\mu_{ij})^m \right)}, \quad \forall j = 1, 2, \dots, c \quad (8)$$

где n - количество точек данных, v_j - представляет j -й центр кластера, m - индекс нечеткости, c - представляет номер центра кластера, μ_{ij} - представляет членство i -го центра данных в j -м центре кластера. d_{ij} - представляет собой евклидово расстояние между i -ми данными и j -м центром кластера .

Основная цель алгоритма нечетких *c*-means- минимизировать

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - v_j\|^2, \quad (9)$$

где $\|x_i - v_j\|$ - евклидово расстояние между этими данными и центром *j*-го кластера.

Пусть $X = \{x_1, x_2, x_3, \dots, x_n\}$ будет набором точек данных, а $V = \{v_1, v_2, v_3, \dots, v_n\}$ будет набором центров.

1. Произвольно выберите центры кластера «*c*».
2. Рассчитайте нечеткое членство μ_{ij} по указанной выше формуле.
3. Вычислите нечеткие центры « v_j », используя (1).
4. Повторяйте шаги 2) и 3) до тех пор, пока не будет достигнуто минимальное значение «*J*» или $\|U(k+1) - U(k)\| < \beta$, где «*k*» - шаг итерации. « β » - критерий завершения между [0, 1]. « $U = (\mu_{ij})_{n \cdot c}$ » - нечеткая матрица принадлежности. «*J*» - целевая функция.

В нашем примере мы хотим показать возможность более быстрого и точного прогнозирования с использованием метода кластеризации с учетом нечеткой логики. Наша задача – спрогнозировать на следующий год спрос на рынке по имеющимся данным. Для этого мы используем MATLAB. Кластеризация числовых данных лежит в основе многих алгоритмов классификации и моделирования систем. Цель кластеризации – выявить естественные группы данных из большого набора данных для создания краткого обзора поведения системы. Нечеткое *c*-средство (FCM) - это метод кластеризации данных, при котором набор данных группируется в *n* кластеров, причем каждый набор данных в наборе данных в определенной степени принадлежит каждому кластеру. Например, определенный объект данных, который находится близко к центру кластера, будет иметь высокую степень принадлежности или принадлежности к этому кластеру, а другой объект данных, который находится далеко от

центра кластера, будет иметь низкую степень принадлежности или членство в нем, чем этот кластер.

Алгоритм действий следующий.

1. Используя команду `fcmdemo`, вы можете запустить графический интерфейс, который позволяет вам опробовать различные настройки параметров для алгоритма нечетких средних и наблюдать их влияние на результирующую двумерную кластеризацию. Вы можете выбрать образец набора данных и произвольное количество кластеров из раскрывающихся меню справа, а затем нажать `Start`, чтобы запустить процесс нечеткой кластеризации `s-meanstr`. Сама кластеризация выполняется функцией `fcm`.

2. Функция `fcm` берет набор данных и желаемое количество кластеров и возвращает оптимальные центры кластеров и оценки членства для каждой точки данных. Вы можете использовать эту информацию для построения нечеткой системы вывода путем создания функций принадлежности, которые представляют нечеткие качества каждого кластера. После этой операции мы получаем следующий рисунок 6.

3. Затем вызовите функцию командной строки `fcm`, чтобы найти два кластера в этом наборе данных, пока целевая функция не перестанет сильно сокращаться.

$[\text{центр}, U, \text{objFcn}] = \text{fcm}(\text{fcmdata}, 2);$

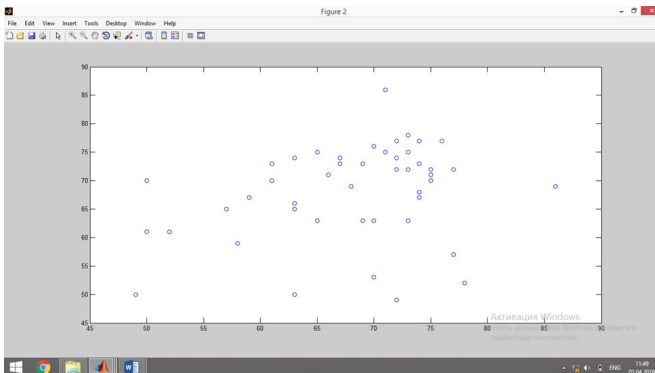


Рисунок 6. Кластерные данные временных рядов за 2 года

Здесь центральная переменная содержит координаты двух центров кластера, U содержит оценки принадлежности для каждой из точек данных, а $objFcm$ содержит историю целевой функции по итерациям.

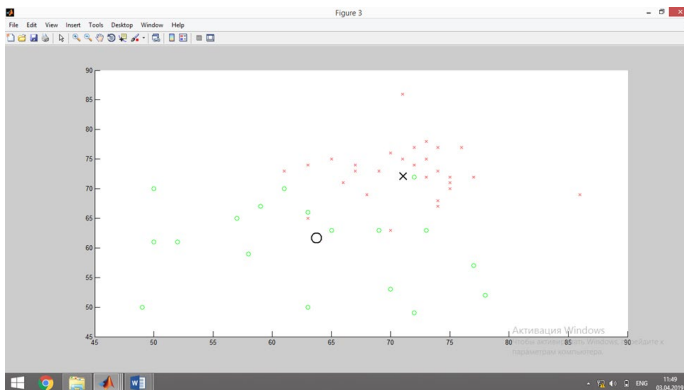


Рисунок 7. Определение центра целевой функции кластеризации

Функция fcm - это итерационный цикл, построенный на основе следующих процедур:

$initfcm$ - инициализирует проблему

$distfcm$ - вычисляет евклидово расстояние

$stepfcm$ - выполняет одну итерацию кластеризации

Чтобы просмотреть прогресс кластеризации, постройте целевую функцию.

Проверка качества обучения отражает способность сетей определять скрытые закономерности в изменениях финансовых показателей и, как следствие, способность адекватно прогнозировать значения коэффициентов финансовой устойчивости. Использование и анализ нечетких временных рядов и метод кластеризации, расширяет возможности систем принятия решений, таким образом, обеспечивает большую точность, чем традиционные методы. Проверка качества обучения отражает способность сетей выявлять скрытые закономерности в изменениях финансовых показателей и, как следствие,

способность адекватно прогнозировать значения коэффициентов финансовой устойчивости.

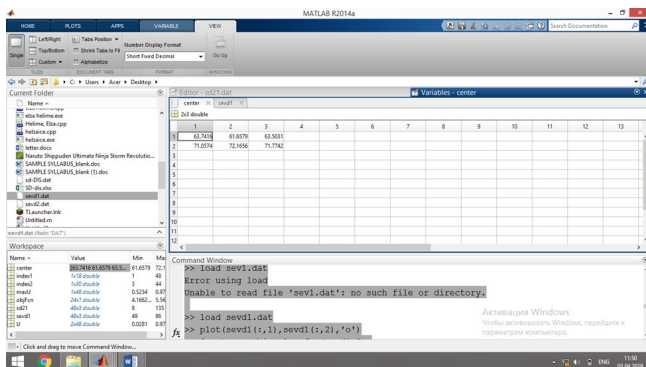


Рисунок 8. Данные, полученные в результате кластеризации

Использование и анализ нечетких временных рядов, использующий метод кластеризации, расширяет возможности систем принятия решений и, таким образом, обеспечивает большую точность, чем традиционные методы. Задача прогноза имеет целью по данным наблюдений предсказать будущие значения измеряемых характеристик изучаемого объекта, т.е. составить прогноз на некоторый отрезок времени. Сейчас разработано и обосновано несколько различных методов прогноза.

Рынок прогнозов – это рынок, на котором люди могут торговать контрактами, которые оплачиваются в зависимости от результатов неизвестных будущих событий. Рыночные цены, генерируемые этими контрактами, можно рассматривать как своего рода коллективное предсказание участников рынка. Эти цены основаны на индивидуальных ожиданиях и готовности инвесторов рисковать своими деньгами ради этих ожиданий.

Прогнозная аналитика часто обсуждается в контексте больших данных. Технические данные, например, поступают от датчиков, инструментов и подключенных систем по всему миру. Данные бизнес-системы в компании могут включать данные транзакций, результаты продаж, жалобы клиентов и

маркетинговую информацию. Все чаще компании принимают решения, основанные на данных, на основе этой ценной информации.

Успешное корпоративное управление зависит от эффективного стратегического и оперативного планирования. Ошибки в планировании часто приводят к огромным расходам, а в некоторых случаях и к убыткам в репутации. Надежные прогнозы вносят важный вклад в эффективное планирование.

Прогнозирование спроса - очень важный фактор для правильного ведения бизнеса. Для компании крайне важно обеспечить эффективное планирование управления операциями, поскольку все организации будут иметь дело с неопределенностью в будущем, следует ожидать некоторой ошибки между прогнозом и фактическим спросом. Цель точного прогноза спроса – минимизировать отклонение между фактическим спросом и прогнозом. Поэтому из-за неопределенности спроса со стороны клиентов они часто сталкиваются с препятствиями, чтобы преодолеть величину и наименьшее количество заказанных товаров. Если возникнет дефицит, это может нанести ущерб компании с точки зрения затрат, в то время как избыточные запасы могут оказать неблагоприятное воздействие на дальнейшие вложения. Это доказывает важность прогнозирования потребительского спроса. Метод нечетких моделей с-means является более подходящим из-за использования точных и неточных данных при анализе спроса. Центры кластеров, полученных методом кластеризации, приведены в “code book” ниже:

Center

$k11mfs = u(1,:); k11mf = k11mfs'; a1 = \max(k11mf);$

$k12mfs = u(2,:); k12mf = k12mfs'; a2 = \max(k12mf);$

$k13mfs = u(3,:); k13mf = k13mfs'; a3 = \max(k13mf).$

Таблица 2.

Нечеткие числа для лингвистических значений при определении спроса на следующий год

Предложение 1-го года	Нечеткое число	Предложение 1-го года	Нечеткое число	Спрогнозированный спрос
66	(59,66,72)	70	(63,70,77)	75
103	(93,103,113)	93	(84,93,102)	65
81	(73,81,89)	86	(78,86,94)	95

На основе этих данных, возможности технологии Data Mining были использованы для изучения прогнозирования потребительского спроса. Используя метод кластеризации FCM, кластеры были получены из вышеуказанных исходных данных, и была построена нечеткая модель. Метод кластеризации был использован для получения знаний из данных. В качестве метода кластеризации был выбран метод нечетких С-средних. В этом методе последовательность вычислений может быть выражена следующим образом:

Алгоритм Fuzzy C-Means пытается минимизировать сумму квадратичных ошибок. Алгоритм основан на итеративной минимизации следующей целевой функции :

$$J(W, C) = \sum_{j=1}^k \sum_{i=1}^n w_{i,j}^p \text{dist}(x_i, c_j)^2 \quad (10)$$

Следующее условие выполняется для общей степени принадлежности данного элемента x_i ко всем кластерам:

$$\sum_{j=1}^k w_{i,j} = 1 \quad (11)$$

В каждом кластере выполняется следующее условие по сумме степеней принадлежности всех элементов:

$$0 < \sum_{i=1}^n w_{i,j} < n \quad (12)$$

Соответствующий центр C_j для кластера c_j определяется следующим образом:

$$c_j = \frac{\sum_{i=1}^n w_{i,j}^p x_i}{\sum_{i=1}^n w_{i,j}^p}. \quad (13)$$

Формула обновления нечеткого разделения может быть получена путем минимизации целевой функции с пределом весов, равным 1:

$$w_{i,j} = \frac{(1 / \text{dist}(x_i, c_j)^2)^{\frac{1}{p-1}}}{\sum_{q=1}^k (1 / \text{dist}(x_i, c_q)^2)^{\frac{1}{p-1}}}. \quad (14)$$

Кластеры были получены указанным выше методом, и была разработана экспертная система.

Включение точек в эти кластеры выглядит следующим образом:

a =

0.7701	0.0915	0.1385
0.4246	0.1192	0.4562
0.2250	0.1828	0.5922
0.0353	0.9145	0.0502
0.2644	0.2167	0.5188
0.2796	0.3273	0.3930
0.1773	0.5273	0.2954
0.3938	0.2950	0.3112
0.3195	0.1567	0.5238
0.2548	0.1662	0.5791
0.2298	0.5131	0.2571
0.2705	0.4564	0.2730
0.7620	0.0801	0.1580
0.9277	0.0263	0.0459
0.8215	0.0513	0.1272
0.7691	0.0498	0.1811
0.1792	0.0561	0.7647
0.0284	0.0180	0.9536
0.0576	0.0723	0.8701
0.1577	0.7158	0.1265

0.4582 0.2956 0.2462
0.7601 0.0973 0.1427

По результатам кластеризации была построена нечеткая модель, состоящая из следующего набора правил.

Если предложения предыдущих двух лет 66 и 70

То спрогнозированный спрос 75

Если предложения предыдущих двух лет 103 и 93

То спрогнозированный спрос 65

Если предложения предыдущих двух лет 81 и 86

То спрогнозированный спрос 95

Причина выбора этого метода в том, что он имеет ряд преимуществ перед другими методами. Например, этот метод больше похож на метод k-средних. Основное отличие состоит в том, что в кластере «Нечеткие-средние» каждая точка имеет определенную степень принадлежности к определенному кластеру, поэтому точка не входит в «кластер» до тех пор, пока у нее слабая Fuzzy-C Means работает медленнее, чем метод K-средних, потому что в этом методе было больше вычислений. Каждая точка оценивается для каждого кластера, и при каждой оценке выполняется больше операций. K-Means основан исключительно на расчетах расстояний, и этот метод не рассчитывает степень аффилированности, т.е. невозможно работать с неточными данными или же существует сильная связь с кластером.

Метод нечетких с-средних является более подходящим, поскольку он использует точные и неточные данные при прогнозе потребительского спроса.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Основные научные результаты, полученные в диссертации, заключаются в следующем:

1. Определен показатель Ляпунова на основе временных данных продаж автомобилей за два года, с последующим доказательством хаотичности системы.

2. Вычислены фрактальная размерность и аттрактор динамической системы в состоянии хаоса.

3. Определена связь между фрактальной размерностью финансовых временных рядов и применением нечеткой логики к определению потребительского спроса продаж автомобилей.

4. Созданы нечеткие модели типа IF-THEN разработанные для прогнозирования спроса.

5. Проведен сравнительный анализ полученных практических результатов прогнозирования с использованием нейронных сетей и с-теанстр.

6. На основе имеющихся данных доказано, что предложенный алгоритм применим при анализе и прогнозировании хаотических временных рядов, с погрешностью 4.33.

7. Научные результаты, полученные в диссертации, могут быть применены к маркетингу в различных сферах. Основная практическая значимость заключается в возможности определения спроса, что в свою очередь позволяет планирование бюджета и снижение затрат.

Основные результаты диссертации опубликованы в следующих работах:

1. Мустафаева С.Р. Нелинейный анализ технологической информации на объектах нефтяной и газовой промышленности / Мустафаева С.Р, Владимирский Э.И. //Оборудование и технологии для нефтегазового комплекса, Москва : - 2012. №1, стр. 48-52.

2. Мустафаева С.Р. Управление хаотическими процессами по данным наблюдений// - Баку : - Вестник Азербайджанской инженерной академии, - 2012. №4 ,(4) стр.57-63.

3. Мустафаева С.Р. Графические методы идентификации и прогнозирования взаимодействующих временных последовательностей // Международная научно – практическая конференция молодых ученых и студентов «Информационные

процессы и технологии, - Севастополь: Изд-во СевНТУ,- 23-27 апреля, - 2012, стр.267.

4. Мустафаева С.Р. Анализ взаимодействующих хаотических временных рядов //Третья Международная научно-практическая конференция «Информационные технологии и компьютерная инженерия», - Винница: Изд-во ВНТУ 29-31 мая 2012 стр.262-263.

5. Мустафаева С.Р. Anomalous diffusion influence of chaotic system by observation // IV International conference “Problems of cybernetics and informatics”,- Baku - , - September 12-14 ,-2012, pp.158-160.

6. Мустафаева С.Р. Нелинейные методы идентификации и прогнозирования временных рядов / Мустафаева С.Р., Мамедов Р.К., Владимирский Э.И. //Автоматизация, телемеханизация и связь в нефтяной промышленности, Москва: - 2012. №2 , - стр.17-21.

7. Мустафаева С.Р. Идентификация и прогнозирование временных рядов // Doktorantların və gənc tədqiqatçıların Azərbaycan xalqının Ümummilli lideri Heydər Əliyevin anadan olmasının 90 iliyinə həsr olunmuş” Azərbaycan 2020: neft-qaz sənayesinin inkişaf perspektivləri” konfrans ,- Bakı: - 2-3 may ,- 2013, s.183-185.

8. Мустафаева С.Р. Нелинейные методы обработки временных рядов //Международная научно-практическая конференция молодых ученых, аспирантов и студентов, - Харьков: Изд-во ,- 2013, стр.82.

9. Мустафаева С.Р. Анализ взаимодействия аномальной диффузии с хаотической системой по наблюдаемому временному ряду / Мустафаева С.Р., Мамедов Р.К., Владимирский Э.И. // Изв. ВТУЗ Азербайджана, - Баку: - 2013. Т. 15, №3 , стр.82-88.

10. Мустафаева С.Р. Аномальная диффузия и ее влияние на хаотические системы / Мустафаева С.Р., Мамедов Р.К., Владимирский Э.И. //Информационные технологии, - Москва: - 2013. №3 ,- стр.15-19.

11. Мустафаева С.Р. Нелинейный рекуррентный анализ временных замеров объектов нефтегазодобычи // - Москва : Автоматизация, телемеханизация и связь в нефтяной промышленности , - 2016. №8, стр.40-42.

12. Мустафаева С.Р. Кросс-рекуррентный анализ коррелированных систем в нефтегазодобыче // - Москва: Автоматизация, телемеханизация и связь в нефтяной промышленности - 2016, №2 стр.3-25.

13. Мустафаева С.Р. Аддитивное и мультипликативное взаимодействие стохастических процессов // - Баку : Вестник Азербайджанской Инженерной Академии - 2017, № 2(9) , стр.115-120.

14. Мустафаева С.Р. Использование метода кластеризации нечеткой логики в вопросах прогнозирования в нефтяной индустрии / Мустафаева С.Р., Н.Г. Джавадов // Автоматизация, телемеханизация и связь в нефтяной промышленности, Москва : - 2019. № 7(552) стр.49-52.

15. Mustafayeva S.R. Calculation of Lyapunov exponent of time series //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-2019 Switzerland , pp.768-771.

16. Mustafayeva S.R. Fractal Dimension Determining for Demand Forecasting //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW- Switzerland , pp.931-937.

17. Мустафаева С.Р.Использование метода кластеризации в автомобильном бизнесе, ADNSU-nun 100 illiyinə həsr olunmuş ADNSU gənc alim və tədqiqatçıların elmi-praktiki konfransı 2020-Bakı: - 2020 ADNSU, стр.452-457.

18. Mustafayeva S.R. Using the clustering method in the automotive business // - Bakı: Azərbaycan Ali Texniki Məktəblərinin Xəbərləri, Bakı: - 2022. s.33-37.

Личный вклад соискателя в трудах, опубликованных в соавторстве:

[2,3,4,5,7,8,11,12,13,15,16,17,18]-самостоятельно проведенные исследования;

[1,6,9] - проведён нелинейный анализ информационных потоков в информационно-измерительных системах;

[4,8,12] - непараметрический анализ хаотических временных рядов;

[15,16,17,18] - предложены исследования в автомобильной области с помощью методов нечеткой логики.

Защита диссертации состоится 16.09.2022 года в 13.00 на заседании Диссертационного совета ЕД 2.02, действующего на базе Азербайджанского Государственного Университета Нефти и Промышленности

Адрес: AZ1010, Баку, проспект Азадлыг, 20. Азербайджанский Государственный Университет Нефти и Промышленности, основное здание, комната 250

С диссертацией можно ознакомиться в библиотеке Азербайджанского Государственного Университета Нефти и Промышленности

Электронная версия диссертации и автореферата размещена на официальном сайте Азербайджанского Государственного Университета Нефти и Промышленности

Автореферат разослан по соответствующим адресам 22.07.2022 года.

Подписано в печать: 30.03.2022

Формат бумаги: А5

Объем: 42259

Тираж: 70